

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 90.3624 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTTIQNGKMSSTIVSE.....ATVLDKNNISSKSTTNPNK 144

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	100.0	773	4	AAB48343
2	748	100.0	2120	3	AAY81710
3	748	100.0	2140	6	ABU01020
4	748	100.0	2140	6	ABU45746
5	748	100.0	2140	8	ADM92113
6	748	100.0	2140	8	ADT50099
7	745	99.6	637	8	ADR94534
8	745	99.6	637	9	AEA58404
9	745	99.6	2138	8	ADK48759
10	615	82.2	117	2	AAW55096
11	615	82.2	117	5	ABP54590
12	615	82.2	117	7	ADC45149
13	110	14.7	746	4	AAG81779
14	110	14.7	778	5	ABP39023
15	110	14.7	778	8	ADS06368
16	108	14.4	188	9	ADZ79639
17	108	14.4	354	9	ADZ72533
18	106.5	14.2	470	8	ADZ56185
19	106.5	14.2	484	3	AAG47777
20	105	14.0	647	9	ADZ79635
21	105	14.0	651	8	AD019012
22	105	14.0	651	8	AD019010
23	104	13.9	665	3	AB018278
24	104	13.9	665	7	AB023606

25	103.5	13.8	169	9	ADZ79634
26	102.5	13.7	775	6	ABU42797
27	101.5	13.6	564	4	ABB61977
28	99.5	13.3	707	6	ABU25018
29	96	12.8	639	9	ADW88472
30	95.5	12.8	635	8	ADS93954
31	95.5	12.8	635	8	ADV83292
32	95.5	12.8	643	8	ADV89902
33	95.5	12.8	643	8	ADV81155
34	95	12.7	645	9	ADW88460
35	95	12.7	645	9	ADW88459
36	95	12.7	645	9	ADW88458
37	94	12.6	645	9	ADW88454
38	94	12.6	645	9	ADW88457
39	94	12.6	645	9	ADW88453
40	94	12.6	645	9	ADW88441
41	94	12.6	645	9	ADW88455
42	94	12.6	645	9	ADW88456
43	94	12.6	645	9	ADW88452
44	93	12.4	645	9	ADW88451
45	93	12.4	1791	8	ADP25441

#### ALIGNMENTS

##### RESULT 1

ID	AAB48343	standard; protein; 773 AA.
XX		
AC	AAB48343;	
XX		
DT	20-APR-2001	(first entry)
XX		
DE	S. pneumoniae	Spl30 polypeptide.
XX		
KW	Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal; bronchial; lung; blood; infection; immune response; immunotherapy; antibacterial; auditory; vaccine.	
XX		
OS	Streptococcus pneumoniae.	
XX		
FN	WO200076540-A2.	
XX		
PD	21-DEC-2000.	
XX		
PF	09-JUN-2000; 2000WO-US015925.	
XX		
PR	10-JUN-1999; 99US-0138453P.	
XX		
PA	(MEDI-) MED IMMUNE INC.	
XX		
PI	Adamou JE, Choi GH;	
XX		
DR	WPI; 2001-112197/12.	
XX		
DR	N-PSDB; AAC84742.	
XX		
PT	New vaccines comprising Spi28 or Spi30 polypeptides, for treating and preventing pneumococcal infections, particularly infections caused by Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or blood infections.	
XX		
PS	Claim 8; Page 51-54; 54pp; English.	
XX		
CC	The invention relates to novel immunogenic polypeptides, Spi28 and Spi30 from S. pneumoniae. Vaccines comprising the polypeptides are useful for the treatment and prevention of pneumococcal infections, particularly infections caused by Streptococcus, such as otitis media, nasopharyngeal, bronchial, lung or blood infections. The antigens are used as immunogenic agents to stimulate an immune response. The antisera and antibodies may also be used in diagnosing and treating pneumococcal infections.	
CC	Recombinant polypeptides serve as a mechanism for stimulating production of antibodies for use in passive immunotherapy, diagnostic reagents, and	

CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 748; DB 4; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 7e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 60  
 DB 630 HRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 689  
 QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 690 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 749  
 QY 121 KDVTATVLDKNNISSKSTTNNPNK 144  
 DB 750 KDVTATVLDKNNISSKSTTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; protein; 2120 AA.  
 XX  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB002452.  
 XX  
 PR 27-JUL-1998; 98GB-00016336.  
 PR 19-MAR-1999; 99US-0125329P.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AA291806.  
 XX  
 PT Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections.  
 XX  
 PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis  
 XX  
 SQ Sequence 2120 AA;

Query Match 100.0%; Score 748; DB 3; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-56;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 60  
 DB 1943 HRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 2002  
 QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 2003 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2062  
 QY 121 KDVTATVLDKNNISSKSTTNNPNK 144  
 DB 2063 KDVTATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 FN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Maignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ASX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 1; SEQ ID NO 1180; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target



CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence and  
CC the second primer is substantially complementary to the target sequence of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
XX Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.9e-66;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLISK 60  
DB 1963 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLISK 2022  
QY 61 DTFIKPVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
DB 2023 DTFIKPVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2082  
QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4  
ABU45746  
ID ABU45746 standard; protein; 2140 AA.  
XX  
XX AC ABU45746;  
XX  
XX DT 19-JUN-2003 (first entry)  
XX  
XX DE Protein encoded by Prokaryotic essential gene #31273.  
XX  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX OS Streptococcus pneumoniae.  
XX  
XX PN WO200277183-A2.  
XX  
XX PD 03-OCT-2002.  
XX  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX  
XX PR 21-MAR-2001; 2001US-00815242.  
XX  
XX PR 06-SEP-2001; 2001US-00948993.  
XX  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX  
XX PR 08-FEB-2002; 2002US-00072851.  
XX  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX PA (ELIT-) ELITRA PHARM INC.  
XX  
XX PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KI, Zykkind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX DR WPI; 2003-029926/02.  
XX  
XX DR N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 73670; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.9e-66;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLISK 60  
DB 1963 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLISK 2022  
QY 61 DTFIKPVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
DB 2023 DTFIKPVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2082  
QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 5  
ADM92113  
ID ADM92113 standard; protein; 2140 AA.  
XX  
XX AC ADM92113;  
XX  
XX DT 03-JUN-2004 (first entry)  
XX  
XX DE S pneumoniae antigenic protein sequence SeqID310.  
XX  
XX KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
XX KW antigenic.  
XX  
XX OS Streptococcus pneumoniae.  
XX

```
PN WO2004020609-A2.
XX 11-MAR-2004.
XX
XX 02-SEP-2003; 2003WO-US027401.
XX
XX 30-AUG-2002; 2002US-0407082P.
XX
XX (TUFT ) UNIV TUFTS.
XX
XX Cam111 A, Hava DL;
XX
XX WPI; 2004-239189/22.
XX N-PSDB; ADM91876.
XX
XX New Streptococcus pneumoniae nucleic acid molecules, useful for
XX diagnosing, treating and preventing active infections of Streptococcus
XX pneumoniae.
XX
XX Claim 27; SEQ ID NO 310; 123pp; English.
XX
XX This invention relates to novel isolated Streptococcus pneumoniae nucleic
XX acid molecules and the antigenic polypeptides encoded by them. The
XX invention may be useful for the production of compounds with an
XX antibacterial activity or for gene therapy. The nucleic acid molecules,
XX compositions and methods disclosed are useful for treating Streptococcus
XX pneumoniae infection. The present sequence is that of an S pneumoniae
XX protein of the invention.
XX
XX Sequence 2140 AA;
SQ
Query Match 100.0%; Score 748; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.9e-66;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 60
DB 1963 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 2022
QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120
DB 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2082
QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
DB 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
RESULT 6
ADT50099
ID ADT50099 standard; protein; 2140 AA.
XX
XX AC ADT50099;
XX
XX 13-JAN-2005 (first entry)
XX
XX S_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.
XX
XX hyperimmune serum reactive antigen; antibacterial; vaccine;
XX bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;
XX sepsis; meningitis.
XX
XX Streptococcus pneumoniae TIGR4.
XX
XX WO2004092209-A2.
XX
XX 28-OCT-2004.
XX
XX 15-APR-2004; 2004WO-EP003984.
XX
XX 15-APR-2003; 2003EP-00450087.
XX
XX (INTE-) INTERCELL AG.
PA
Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;
WPI; 2004-758335/74.
N-PSDB; ADT49955.
New hyperimmune serum reactive antigens from Streptococcus pneumoniae,
and encoding nucleic acid molecules, useful for diagnosing, preventing or
treating S. pneumoniae infections.
Disclosure; SEQ ID NO 177; 191pp; English.
This invention relates to novel nucleic acids encoding hyperimmune serum
reactive antigens, or fragments derived thereof. Specifically, it refers
to antigens selected from peptides and serum reactive epitopes that can
be used in pharmaceutical compositions that exhibit antibacterial
activity. The present invention describes a composition (including the
nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)
that is useful for manufacturing a medicament such as a vaccine, which
can be used to treat or prevent bacterial infections, particularly S.
pneumoniae infections that cause pharyngitis, otitis media, pneumonia,
bacteraemia sepsis and meningitis. The antigen or its fragment may also
be used for isolating, purifying and/ or identifying an interaction
partner of the hyperimmune serum reactive antigen, as well as for
manufacturing a functional nucleic acid selected from aptamers and
spiegelmers or for manufacturing a functional ribonucleic acid selected
from ribozymes, antisense nucleic acids and siRNA. This polypeptide
sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen
of the invention.
Sequence 2140 AA;
SQ
Query Match 100.0%; Score 748; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.9e-66;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 60
DB 1963 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 2022
QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120
DB 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2082
QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
DB 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
RESULT 7
ADR94534
ID ADR94534 standard; protein; 637 AA.
XX
XX AC ADR94534;
XX
XX 16-DEC-2004 (first entry)
XX
XX Novel S. pneumoniae protein sequence, SEQ ID 3169.
XX
XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
XX bacterial infection.
XX
XX Streptococcus pneumoniae.
XX
XX US6800744-B1.
XX
XX 05-OCT-2004.
XX
XX 30-JUN-1998; 98US-00107433.
XX
XX 02-JUL-1997; 97US-0051553P.
XX
XX 12-MAY-1998; 98US-0085131P.
XX
```

PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm LA, Bush D;  
 XX WPI; 2004-697205/68.  
 DR N-PSDB; ADR91931.  
 XX  
 DR New isolated nucleic acid encoding a Streptococcus pneumoniae  
 PT polypeptide, useful for diagnosing, preventing and/or treating  
 PT pathological conditions resulting from the bacterial infection.  
 XX  
 PS Disclosure; SEQ ID NO 3169; 151pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising a sequence  
 CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
 CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
 CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95692,  
 CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
 CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
 CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
 CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
 CC hybridizable under high stringency conditions to the nucleotide sequence.  
 CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
 CC Also included are a recombinant expression vector comprising the isolated  
 CC nucleic acid cited above operably linked to a transcription regulatory  
 CC element, a cell comprising the recombinant expression vector and a probe  
 CC comprising at least 20 consecutive nucleotides of the nucleotide  
 CC sequences as cited above. The methods and compositions of the present  
 CC invention are useful for the diagnosis, prevention and/or treatment of  
 CC pathological conditions resulting from bacterial infection by  
 CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
 CC otitis media. The present sequence is one of the 2603 disclosed S.  
 CC pneumoniae protein sequences. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
 XX  
 SQ Sequence 637 AA;  
 Query Match 99.6%; Score 745; DB 8; Length 637;  
 Best Local Similarity 99.3%; Pred. No. 1.1e-66;  
 Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWISGFGKDGAGVYINLSK 60  
 Db 460 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWISGFGKDGAGVYINLSK 519  
 QY 61 DTFFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQREHSQKSDST 120  
 Db 520 DTFFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQREHSQKSDST 579  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 Db 580 KDVTATVLDKNNISSKSTNNPNK 603  
 RESULT 8  
 AEA58404  
 ID AEA58404 standard; protein; 637 AA.  
 XX AEA58404;  
 AC AEA58404;  
 XX  
 XX 25-AUG-2005 (first entry)  
 XX  
 XX Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
 XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
 KW vaccine.  
 XX  
 XX Streptococcus pneumoniae.  
 OS  
 XX US2005136404-A1.  
 FN  
 XX

PD 23-JUN-2005.  
 XX 10-JUL-2003; 2003US-00617320.  
 PF  
 XX 02-JUL-1997; 97US-0051553P.  
 PR 12-MAY-1998; 98US-0085131P.  
 PR 30-JUN-1998; 98US-00107433.  
 XX  
 XX (DOUC/) DOUCETTE-STAMM L A.  
 PA (BUSH/) BUSH D.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 XX WPI; 2005-477576/48.  
 DR N-PSDB; AEA55801.  
 DR  
 XX New isolated nucleic acid molecules and encoded polypeptides useful for  
 PT diagnosing, preventing or treating bacterial infections, particularly  
 PT Streptococcus pneumoniae infection.  
 XX  
 XX Claim 5; SEQ ID NO 3169; 144pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule for detecting,  
 CC preventing or treating pathological conditions resulting from bacterial  
 CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
 CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
 CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
 CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
 CC sequence of at least 8 nucleotides in length, where the sequence is  
 CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
 CC (a). Also described: (1) a recombinant expression vector comprising the  
 CC above nucleic acid operably linked to a transcription regulatory element;  
 CC (2) a cell comprising the recombinant expression vector; (3) producing an  
 CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
 CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
 CC treating a subject for S. pneumoniae infection; (6) a recombinant or its  
 CC substantially pure preparation of an S. pneumoniae polypeptide or its  
 CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
 CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
 CC infection, comprising an amount of the above nucleic acid or polypeptide;  
 CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
 CC (9) a computer readable medium having recorded the nucleotide sequences  
 CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
 CC fragments of the Streptococcus genome of commercial importance. The  
 CC composition and methods are useful for diagnosing, preventing or treating  
 CC bacterial infections, particularly S. pneumoniae infection. The present  
 CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
 CC present invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from the USPTO web site.  
 XX  
 SQ Sequence 637 AA;  
 Query Match 99.6%; Score 745; DB 9; Length 637;  
 Best Local Similarity 99.3%; Pred. No. 1.1e-66;  
 Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWISGFGKDGAGVYINLSK 60  
 Db 460 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWISGFGKDGAGVYINLSK 519  
 QY 61 DTFFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQREHSQKSDST 120  
 Db 520 DTFFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQREHSQKSDST 579  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 Db 580 KDVTATVLDKNNISSKSTNNPNK 603  
 RESULT 9  
 ADK48759  
 ID ADK48759 standard; protein; 2138 AA.

[illegible]

XX	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW	detection; pneumonia; otitis media; meningitis.
XX	Streptococcus pneumoniae.
OS	Streptococcus pneumoniae.
XX	WO9818930-A2.
PN	07-MAY-1998.
XX	30-OCT-1997; 97WO-USO19422..
PF	31-OCT-1996; 96US-0029960P.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI	WFI; 1998-272224/24.
DR	N-PSDB; AAV27357.
XX	Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT	- or their epitope-containing fragments, useful in protective or
PT	therapeutic vaccines, and for diagnosis.
XX	Claim 11; Page 62; 118pp; English.
PS	The present sequence represents a protein from Streptococcus pneumoniae.
XX	CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC	can be useful in vaccines for inducing protective antibodies against
CC	Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC	pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC	are used to detect Streptococcus infection (by usual hybridisation or
CC	amplification methods), also for isolating Streptococcus genes or their
CC	allelic variants. The protein can be used similarly to detect specific
CC	antibodies in standard immunoassays, especially for diagnosing or
CC	monitoring infections. Antibodies which bind the protein are used to
CC	detect corresponding antigens, to purify the protein and for passive
CC	immunisation (optionally coupled to a toxin). Vaccines are administered,
CC	e.g. by injection, orally or through the skin, typically at 0.01-1000
CC	(especially 10-300) mu g/ml per dose
XX	Sequence 117 AA;
SQ	Query Match 82.2%; Score 615; DB 2; Length 117;
	Best Local Similarity 100.0%; Pred. No. 1.8e-54;
	Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	28 YKGELEKGYPFGWEISGPEGKKDAGYVNLNLSKDTFKPVFKIBEKKEENKPTFDVSK 87
Dd	1 YKGELEKGYPFGWEISGPEGKKDAGYVNLNLSKDTFKPVFKIBEKKEENKPTFDVSK 60
Qy	88 KKNQPQNWSQLNESHKEDLQREHSQKSDSTKDVTATVLDDKNNSKSTTNNPK 144
Dd	61 KKNQPQNWSQLNESHKEDLQREHSQKSDSTKDVTATVLDDKNNSKSTTNNPK 117
RESULT 11	
ABP54590	
ID	ABP54590 standard; protein; 117 AA.
XX	ABP54590;
AC	
XX	04-SEP-2002 (first entry)
DT	
XX	S. pneumoniae SP043 protein sequence SEQ ID NO:68.
DE	Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX	antibacterial; Streptococcal infection; detection.
KW	Streptococcus pneumoniae.
XX	US2002061545-A1.
OS	
XX	
FN	

```

XX PD 23-MAY-2002.
XX PF
XX PA 22-JAN-2001; 2001US-00765272.
XX PR
XX PI 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
Query Match 82.2%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.8e-54;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKEENKPTFDVSK 87
DB 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKEENKPTFDVSK 60
QY 88 KDNFQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144
DB 61 KDNFQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX US6573082-B1.
XX PN 03-JUN-2003.
XX PD 28-MAR-2000; 2000US-00536784.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR

```

```

PR 30-OCT-1997; 97US-00961083.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 58pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
Query Match 82.2%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.8e-54;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKEENKPTFDVSK 87
DB 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKEENKPTFDVSK 60
QY 88 KDNFQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144
DB 61 KDNFQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
RESULT 13
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX AC AAG81779;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX KW endocarditis.
XX OS Staphylococcus epidermidis.
XX XX WO200134809-A2.
XX PN 17-MAY-2001.
XX PD 09-NOV-2000; 2000WO-US030782.
XX PF 09-NOV-1999; 99US-0164258P.
XX PR (GLAX ) GLAXO GROUP LTD.
XX PA Kimmerly WJ;
XX PI WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.
XX DR

```

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
PS Claim 18; Page 208; 2189pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 746 AA;

Query Match 14.7%; Score 110; DB 4; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.062;  
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
QY 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGFE-----G 48  
Db 596 ITINGKQIKQOSVKSQTKVLPHSKVLMTDGELTMP-DMTGWTKEVDLAFEDLTJKLVS 654  
QY 49 KKDAGYVIN--LSKDTFIKPVFKKIEEKEENKPTFDVS---KKKNPQVNHSQLNES 102  
Db 655 TKGNGFVTNQISKGQIILK-----NKKIEVLSAEDTDDQKTDDESDN 701  
QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
Db 702 KSKKQKADHDHNTSSSTKN-----DKSNADSKNDSDD 734

RESULT 14  
ABP39023  
ID ABP39023 standard; protein; 778 AA.  
XX  
AC ABP39023;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-00134001.  
XX  
PR 14-AUG-1997; 97US-0055779P.  
PR 08-NOV-1997; 97US-0064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.  
DR N-PSDB; ABN91568.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis  
PT polypeptide, useful for diagnosing and treating bacterial infections.  
PS Disclosure; SEQ ID NO 3868; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences can  
CC also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life cycle  
CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the USPTO web site  
XX  
SQ Sequence 778 AA;

Query Match 14.7%; Score 110; DB 5; Length 778;  
Best Local Similarity 27.0%; Pred. No. 0.066;  
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
QY 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGFE-----G 48  
Db 628 ITINGKQIKQOSVKSQTKVLPHSKVLMTDGELTMP-DMTGWTKEVDLAFEDLTJKLVS 686  
QY 49 KKDAGYVIN--LSKDTFIKPVFKKIEEKEENKPTFDVS---KKKNPQVNHSQLNES 102  
Db 687 TKGNGFVTNQISKGQIILK-----NKKIEVLSAEDTDDQKTDDESDN 733  
QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
Db 734 KSKKQKADHDHNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 15  
ADS06368  
ID ADS06368 standard; protein; 778 AA.  
XX  
AC ADS06368;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Staphylococcus epidermis polypeptide seqid 5663.  
XX  
KW antibacterial; vaccine; antiseptic therapy; Staphylococcus epidermidis;  
KW recombinant expression vector; infection; computer readable medium;  
KW computer based system.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US2004147734-A1.  
XX  
PD 29-JUL-2004.  
XX  
PF 01-DEC-2003; 2003US-00724972.  
XX  
PR 08-NOV-1997; 97US-0064964P.  
PR 13-AUG-1998; 98US-00134001.  
PR 29-NOV-1999; 99US-00450969.  
XX  
PA (DOUC/) DOUCETTE-STAMM L.  
PA (BUSH/) BUSH D.  
XX  
PI Doucette-Stamm L, Bush D;  
XX  
DR WPI; 2004-580138/56.  
DR N-PSDB; ADS02596.  
XX





PD 27-APR-2005.  
XX  
PF 24-OCT-2003; 2003EP-00292673.  
XX  
XX 24-OCT-2003; 2003EP-00292673.  
PR  
XX (INSP ) INST PASTEUR.  
PA  
XX Drulhe P;  
PI  
XX WPI; 2005-323987/34.  
DR  
DR N-PSDB; AD272252.  
XX  
XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
PT falciparum, which encode proteins useful for preparing vaccine  
PT compositions against malaria.  
XX  
XX Disclosure; SEQ ID NO 2; 137pp; English.  
XX  
XX The present invention relates to the protection against malaria. More  
CC particularly, the invention pertains to a family of MSP-3 (merozoite  
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
CC falciparum, highly conserved in P. falciparum strains, simultaneously  
CC expressed in P. falciparum at the erythrocytic stages and encoding  
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
CC their N-terminal extremity and which are located at the merozoite  
CC surface. The characterization of this gene family enables the definition  
CC of immunogenic and vaccine compositions against P. falciparum. The  
CC present sequence is the P. falciparum MSP-3-1 protein.  
XX  
SQ Sequence 354 AA;  
Query Match 14.4%; Score 108; DB 9; Length 354;  
Best Local Similarity 23.2%; Pred. No. 0.035;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 7 IQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGP--EGKKDAG-----VVINLS 59  
DB 181 VLKAEAS-----YDYL-----GWFGGVPFHKKENMLSHLYVSKD 221  
QY 60 KDTFKVPKIEEKKE-----ENKPTFDVSKKQNPQVNHSLQNSHRKE 106  
DB 222 KENISKENDVDLKEEAEETEEBLEKNEEETESISEDEEBEKEEENKXK 281  
QY 107 DLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
DB 282 EOEKEQSNENNDQKDMA-----QNLISKNNNN 311  
RESULT 18  
ADTS6185  
ID ADTS6185 standard; protein; 470 AA.  
XX  
XX ADTS6185;  
XX  
XX 13-JAN-2005 (first entry)  
DT  
XX Plant polypeptide, SEQ ID 6262.  
DE  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
OS Viridiplantae.  
XX  
XX US2004216190-A1.  
PN  
XX 28-OCT-2004.  
PD  
XX 18-DEC-2003; 2003US-00739930.  
PF

XX 28-APR-2003; 2003US-00424599.  
PR  
XX 28-APR-2003; 2003US-00425115.  
XX  
XX (KOVA/) KOVALIC D K.  
PA  
XX Kovalic DK;  
PI  
XX WPI; 2004-757369/74.  
DR  
XX  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX  
XX Claim 2; SEQ ID NO 6262; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
SQ Sequence 470 AA;  
Query Match 14.2%; Score 106.5; DB 8; Length 470;  
Best Local Similarity 20.3%; Pred. No. 0.074;  
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;  
QY 1 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFE-----GKKDA 52  
DB 84 NRVTDTVQNNNGESK-----YVDLARRIRYDE-EATGSQSAQRIDHPNQKNV 131  
QY 53 GYVINLSKDTPIKVPFKKIEEKEENKPTPDVSKKQDN-----PQ 93  
DB 132 GITEKAFENSPIETSHRVDDNKRINNQNKNFTAAKSENNAVSRVSFGADHKRAEYMGKPM 191  
QY 94 VNHSLQNE-----SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 144  
DB 192 ENRDQVRQTSAEKSHRKNVTKSEKPRDQGVKKTEAKOKDRNKEKKEKTESINK 249

RESULT 19	PR	18-JUN-1999;	99US-0139459P;
AAG47777	PR	18-JUN-1999;	99US-0139460P;
ID	PR	18-JUN-1999;	99US-0139461P;
AAAG47777 standard; protein; 484 AA.	PR	18-JUN-1999;	99US-0139462P;
AC	PR	18-JUN-1999;	99US-0139463P;
AC	PR	18-JUN-1999;	99US-0139750P;
XX	PR	18-JUN-1999;	99US-0139763P;
DT	PR	18-JUN-1999;	99US-0139899P;
DE	PR	21-JUN-1999;	99US-0139817P;
XX	PR	22-JUN-1999;	99US-0139899P;
XX	PR	23-JUN-1999;	99US-0140353P;
XX	PR	23-JUN-1999;	99US-0140354P;
KW	PR	24-JUN-1999;	99US-0140695P;
KW	PR	28-JUN-1999;	99US-0140823P;
KW	PR	29-JUN-1999;	99US-0140991P;
XX	PR	30-JUN-1999;	99US-0141287P;
OS	PR	01-JUL-1999;	99US-0141842P;
XX	PR	01-JUL-1999;	99US-0142154P;
XX	PR	02-JUL-1999;	99US-0142055P;
XX	PR	06-JUL-1999;	99US-01423900P;
XX	PR	08-JUL-1999;	99US-0142803P;
XX	PR	09-JUL-1999;	99US-0142920P;
XX	PR	12-JUL-1999;	99US-0142977P;
XX	PR	13-JUL-1999;	99US-0143542P;
XX	PR	14-JUL-1999;	99US-0143624P;
XX	PR	15-JUL-1999;	99US-0144005P;
XX	PR	16-JUL-1999;	99US-0144085P;
XX	PR	16-JUL-1999;	99US-0144086P;
XX	PR	19-JUL-1999;	99US-0144325P;
XX	PR	19-JUL-1999;	99US-0144331P;
XX	PR	19-JUL-1999;	99US-0144332P;
XX	PR	19-JUL-1999;	99US-0144333P;
XX	PR	19-JUL-1999;	99US-0144334P;
XX	PR	19-JUL-1999;	99US-0144335P;
XX	PR	20-JUL-1999;	99US-0144352P;
XX	PR	20-JUL-1999;	99US-0144632P;
XX	PR	20-JUL-1999;	99US-0144884P;
XX	PR	21-JUL-1999;	99US-0144814P;
XX	PR	21-JUL-1999;	99US-0145086P;
XX	PR	21-JUL-1999;	99US-0145088P;
XX	PR	22-JUL-1999;	99US-0145085P;
XX	PR	22-JUL-1999;	99US-0145087P;
XX	PR	22-JUL-1999;	99US-0145089P;
XX	PR	22-JUL-1999;	99US-0145192P;
XX	PR	23-JUL-1999;	99US-0145145P;
XX	PR	23-JUL-1999;	99US-0145218P;
XX	PR	23-JUL-1999;	99US-0145224P;
XX	PR	26-JUL-1999;	99US-0145276P;
XX	PR	27-JUL-1999;	99US-0145913P;
XX	PR	27-JUL-1999;	99US-0145918P;
XX	PR	27-JUL-1999;	99US-0145919P;
XX	PR	28-JUL-1999;	99US-0145951P;
XX	PR	02-AUG-1999;	99US-0146386P;
XX	PR	02-AUG-1999;	99US-0146388P;
XX	PR	02-AUG-1999;	99US-0146389P;
XX	PR	03-AUG-1999;	99US-0147038P;
XX	PR	04-AUG-1999;	99US-0147204P;
XX	PR	04-AUG-1999;	99US-0147302P;
XX	PR	05-AUG-1999;	99US-0147192P;
XX	PR	05-AUG-1999;	99US-0147260P;
XX	PR	06-AUG-1999;	99US-0147303P;
XX	PR	06-AUG-1999;	99US-0147416P;
XX	PR	09-AUG-1999;	99US-0147933P;
XX	PR	09-AUG-1999;	99US-0147935P;
XX	PR	10-AUG-1999;	99US-0148171P;
XX	PR	11-AUG-1999;	99US-0148319P;
XX	PR	12-AUG-1999;	99US-0148341P;
XX	PR	13-AUG-1999;	99US-0148565P;
XX	PR	13-AUG-1999;	99US-0148684P;
XX	PR	16-AUG-1999;	99US-0149368P;
XX	PR	17-AUG-1999;	99US-0149317P;
XX	PR	18-AUG-1999;	99US-0149426P;
XX	PR	20-AUG-1999;	99US-0149722P;
XX	PR	20-AUG-1999;	99US-0149723P;
XX	PR	20-AUG-1999;	99US-0149725P;
XX	PR	20-AUG-1999;	99US-0139458P;
XX	PR	20-AUG-1999;	99US-0139457P;
XX	PR	20-AUG-1999;	99US-0139456P;
XX	PR	20-AUG-1999;	99US-0139455P;
XX	PR	20-AUG-1999;	99US-0139454P;
XX	PR	20-AUG-1999;	99US-0139453P;
XX	PR	20-AUG-1999;	99US-0139452P;
XX	PR	20-AUG-1999;	99US-0139451P;
XX	PR	20-AUG-1999;	99US-0138847P;
XX	PR	20-AUG-1999;	99US-0138540P;
XX	PR	20-AUG-1999;	99US-0138094P;
XX	PR	20-AUG-1999;	







QY 58 LSKDTFIKVPFKIEKKK-----EKNKTPD-----VSKKKNPQVNHSLNESH 103  
Db 175 -----VKGVKLEKKKBEKISDDHKVKEENKSDDHKVENKSDDHKVENKSDDH 228  
QY 104 RKEDLQR-BEHSKSDSTKDTATVLDKNNISKSTNNPNK 144  
Db 229 KLEVKVVEHEDEEE-----DKCKESENKQNDENK 261

## RESULT 25

ID ADZ79634 standard; protein; 169 AA.

XX AC ADZ79634;

XX DT 14-JUL-2005 (first entry)

XX DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.

XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
XX immunotherapy; malaria; antimalarial; vaccine.

XX OS Plasmodium falciparum.

XX FH Key Location/Qualifiers

XX FT Region 1..169

XX FT /note="Amino acid residues 212-380 of MSP3"

XX PN WO2005040206-A1.

XX PD 06-MAY-2005.

XX PF 22-OCT-2004; 2004WO-BP012910.

XX PR 24-OCT-2003; 2003US-00691672.

XX PA (INSP ) INST PASTEUR.

XX PI Druilhe P;

XX DR WPI; 2005-355821/36.

XX PT Chimeric molecule useful for preparing vaccine composition against  
PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.

XX PS Claim 2; SEQ ID NO 2; 79pp; English.

XX CC The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3  
CC protein (amino acid residues 212-380).

XX SQ Sequence 169 AA;

Query Match 13.8%; Score 103.5; DB 9; Length 169;

Best Local Similarity 25.2%; Pred. No. 0.037;  
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
QY 11 KEMSTIVSEEDFILPVYKGLGKGYQFDGWEISGF--EGKKDAG-----YYINLSKDTF 63  
Db 1 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSDKKENI 41  
QY 64 IKPVFKKIEKKBEENKFTPDVSKKKNPQVNHSLN-----ESHRKEDLQR 110  
Db 42 SKENDDVLDG-KBEAEABTEEELEKNEVEETESISEDEEEBEKEEENKKEKQEK 100  
QY 111 EHSKSDSTKDTATVLDKNNISKSTNN 141  
Db 101 EQSNENNDDKKDMEA-----QNLISKNQNNN 126

## RESULT 26

ID ABU42797

XX ABU42797 standard; protein; 775 AA.

XX AC ABU42797;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #28324.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus epidermidis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX W Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA46667.

XX CC New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 70721; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 775 AA;

Query Match 13.7%; Score 102.5; DB 6; Length 775;  
Best Local Similarity 24.7%; Pred. No. 0.38;  
Matches 38; Conservative 26; Mismatches 61; Indels 29; Gaps 7;  
QY 5 VTIQKEMSTIVSEEDFILPVYK-----GELEKGYQFDGW-----EISGPF-----G 48  
Db 625 ITIGNGKQIKQOSVKSGTKVLPKSHKVMMLTDGELTMP-DMTGWTKEVDLAFEDLTKIKVS 683  
QY 49 KKDAGYVIN--LSKDTFFIKPVFK-----KIEEKGEEENKPTFDVSKKDNPOVNHQS 98  
Db 684 TKGNGFVTNQSIKSGQIIKKNKDKIEVLSAETDDQDKTDESDSKKSKKDKVDNNSN 743  
QY 99 LNESHRKEDLQREHSQKSDSTKDVATVLDKNN 132  
Db 744 ASSSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 27  
ID ABB61977  
AC ABB61977; protein; 564 AA.  
XX  
XX ABB61977;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 12723.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL06080.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 564 AA;

Query Match 13.6%; Score 101.5; DB 4; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.31;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 21 EDFILPVYKGELEKGYQFDGW-----EISGFEKKGAGYVI-----NLSKDTFIK 65  
Db 78 EDLDTPLSESRFSK--VFDGWVDEHDEHDGHVQBPSEALDDHDEHDDHDEHDEDEE 135  
QY 66 PVFKKIEEKEENKPT-----FDVSKKDNPOVNHSQLNESHRKEDLQREHSQKSDS 119  
Db 136 PLTELEEEEEEPEEPTDEDEPAADNEYEDEENNA--GENITAEAESEEEEDND 193  
QY 120 TKDVTATVLDKNNISSKST 138  
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 28  
ID ABU25018  
AC ABU25018; standard; protein; 707 AA.  
XX  
XX ABU25018;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by prokaryotic essential gene #10545.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Clostridium difficile.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA28888.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 52942; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense





XX WPI; 2004-357201/33.  
DR New nucleic acid molecules and encoded adhesion factors and/or fibrinogen  
XX -binding polypeptides for diagnosing, preventing or treating bacterial  
PT infections, preferably Streptococcus agalactiae infection.  
PT Claim 13; SEQ ID NO 19; 225pp; English.  
PS  
XX  
CC The invention relates to a novel isolated nucleic acid molecule encoding  
CC a fibrinogen-binding polypeptide or its fragment, or an adhesion factor  
CC or its fragment. The invention further comprises: a vector comprising the  
CC above nucleic acid molecule; a cell, preferably a host cell, comprising  
CC the vector; a polypeptide, preferably a fibrinogen-binding polypeptide  
CC and/or an adhesion factor, comprising an amino acid sequence encoded by  
CC the above nucleic acid molecule; a process for producing the above  
CC polypeptide or its fragment; a process for producing a cell that  
CC expresses the polypeptide or its fragment; a pharmaceutical composition,  
CC especially a vaccine, comprising the polypeptide or its fragment, or the  
CC above nucleic acid molecule; an antibody, or its part, that binds to at  
CC least a selective part of the polypeptide or its fragment; methods for  
CC identifying an antagonist capable of reducing or inhibiting the activity  
CC of the polypeptide or its fragment, capable of binding to the  
CC polypeptide, or capable of reducing or inhibiting the interaction  
CC activity of the polypeptide or its fragment to its interaction partner;  
CC an antagonist identified by the above method; processes for in vitro  
CC diagnosis of a bacterial infection, preferably Streptococcus agalactiae  
CC infection, or a disease related to the expression of the above  
CC polypeptide or its fragment; and an affinity device comprising a support  
CC material and immobilized to the support material the above polypeptide or  
CC nucleic acid molecule. The fibrinogen-binding polypeptide has  
CC antibacterial activity. The nucleic acids may be used in gene therapy to  
CC treat disorders. The nucleic acid is used for manufacturing or generating  
CC functional ribonucleic acids selected from ribozymes, antisense nucleic  
CC acids and siRNA. The polypeptide or its fragment is useful for  
CC manufacturing a medicament, especially a vaccine against bacterial  
CC infection. The polypeptide is also used for manufacturing an antibody  
CC that may be used for manufacturing a medicament for treating or  
CC preventing bacterial infections, especially S. agalactiae infections. The  
CC polypeptide is also used as an antigen, or for isolating, purifying  
CC and/or identifying an interaction partner of the polypeptide or its  
CC fragment, or for generating a peptide binding to the polypeptide, such as  
CC anticalines. It may also be used for manufacturing or generating a  
CC functional nucleic acid selected from aptamers and Spiegelmers. This  
CC sequence represents a fibrinogen-binding polypeptide of the invention.  
XX  
SQ Sequence 635 AA;  
Query Match 12.8%; Score 95.5; DB 8; Length 635;  
Best Local Similarity 24.5%; Pred. No. 1.5;  
Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
QY 9 NGKMSSTIVSEEDFILP--VYKGBL-----EKGYQFD-----GWEISGFEGKKDAGYVI 56  
Db 313 NTEPLTSLYLNKKEFLVPNIPYKNKILREEDKYSFDEDEEFGNELLSYNKLKNEVLVP 372  
QY 57 NLSKDTPIKVPFKLIEKKGEENKPTFDVSKCKDNPQVNHSHKEDLQREHS-- 114  
Db 373 NITTTSLKLP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPT 425  
QY 115 QKSDSTKDVATVLDKN--NISKSTT 139  
Db 426 PQKSTVKEQTEQKVSNGTQVEKKSET 452  
RESULT 31  
ADV83292  
ID ADV83292 standard; protein; 635 AA.  
XX  
AC ADV83292;  
XX  
DT 24-FEB-2005 (first entry)  
XX

DE Streptococcus agalactiae protein, SEQ ID 4433.  
XX  
KW Antibacterial; vaccine; bacterial infection.  
XX  
OS Streptococcus agalactiae.  
XX  
PN WO200292818-A2.  
XX  
PD 21-NOV-2002.  
XX  
PF 26-APR-2002; 2002WO-IB003059.  
XX  
PR 26-APR-2001; 2001FR-00005642.  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;  
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX  
DR WPI; 2004-101891/11.  
XX  
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
PT and identification of therapeutic targets.  
XX  
PS Claim 6; SEQ ID NO 4433; 439pp; French.  
XX  
CC The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
SQ Sequence 635 AA;  
Query Match 12.8%; Score 95.5; DB 8; Length 635;  
Best Local Similarity 24.5%; Pred. No. 1.5;  
Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
QY 9 NGKMSSTIVSEEDFILP--VYKGBL-----EKGYQFD-----GWEISGFEGKKDAGYVI 56  
Db 313 NTEPLTSLYLNKKEFLVPNIPYKNKILREEDKYSFDEDEEFGNELLSYNKLKNEVLVP 372  
QY 57 NLSKDTPIKVPFKLIEKKGEENKPTFDVSKCKDNPQVNHSHKEDLQREHS-- 114  
Db 373 NITTTSLKLP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPT 425  
QY 115 QKSDSTKDVATVLDKN--NISKSTT 139  
Db 426 PQKSTVKEQTEQKVSNGTQVEKKSET 452  
RESULT 32  
ADV89902  
ID ADV89902 standard; protein; 643 AA.  
XX  
AC ADV89902;  
XX  
DT 24-FEB-2005 (first entry)  
XX

DE Streptococcus agalactiae protein sequence, SEQ ID 2296.

KW Antibacterial; Vaccine; bacterial infection.

OS Streptococcus agalactiae.

PN FR2824074-A1.

XX 31-OCT-2002.

XX 26-APR-2001; 2001FR-00005642.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

PI Glaser P, Ruanio C, Chevalier P, Frangeul L, Lalioui L;

PI Zouine M, Couve B, Buchrieser C, Poyart C, Trieu CP, Kunst P;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

PS Claim 6; SEQ ID NO 2296; 2687bp; French.

XX The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II; ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation to medicines and/or analogues, functions related to transporters, cell membrane proteins and cofactors, prosthetic groups and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids.

XX Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is equivalent for the present basic patent FR2824074A1. WO200292818A2 contains 6617 sequence whereas the present patent only contains 2344 sequences.

XX Sequence 643 AA;

Query Match 12.8%; Score 95.5; DB 8; Length 643;

Best Local Similarity 24.5%; Pred. No. 1.5;

Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;

QY 9 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGPEGKDGVI 56

Db 321 NTEPLTSLYLNKKEKFLPNIPYKNGKILRREBDKYSFDEDEEFGNLLSYNKLKNEVL 380

QY 57 NLSKDTFTKPVFKKLEKKEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREHS-- 114

Db 381 NITSTILKP-----PEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPT 433

QY 115 QKSDSTKDVATVLDKN--NISKSTT 139

Db 434 PQKTVKEQTEQKVSQNTQVEKKSET 460

RESULT 33

ADV81155

XX ADV81155 standard; protein; 643 AA.

XX AC ADV81155;

XX 24-FEB-2005 (first entry)

XX

DE Streptococcus agalactiae protein, SEQ ID 2296.

KW Antibacterial; vaccine; bacterial infection.

OS Streptococcus agalactiae.

XX WO200292818-A2.

XX 21-NOV-2002.

XX 26-APR-2002; 2002WO-IB003059.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

PI Glaser P, Ruanio C, Chevalier P, Frangeul L, Lalioui L;

PI Zouine M, Couve B, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst P;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

PS Claim 6; SEQ ID NO 2296; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation to medicines and/or analogues, functions related to transporters, cell membrane proteins and cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. The complete genome of Streptococcus agalactiae is given in ADV81204. Note: The present patent is an equivalent for the basic patent FR2824074A1, which contains only 2344 sequences.

XX Sequence 643 AA;

Query Match 12.8%; Score 95.5; DB 8; Length 643;

Best Local Similarity 24.5%; Pred. No. 1.5;

Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;

QY 9 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGPEGKDGVI 56

Db 321 NTEPLTSLYLNKKEKFLPNIPYKNGKILRREBDKYSFDEDEEFGNLLSYNKLKNEVL 380

QY 57 NLSKDTFTKPVFKKLEKKEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREHS-- 114

Db 381 NITSTILKP-----PEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPT 433

QY 115 QKSDSTKDVATVLDKN--NISKSTT 139

Db 434 PQKTVKEQTEQKVSQNTQVEKKSET 460

RESULT 34...

ADV88460

XX ADV88460 standard; protein; 645 AA.

XX AC ADV88460;

XX 21-APR-2005 (first entry)

XX











**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 93.5369 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVTQNG.....ATVLDKXNISKSTTNPKNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	799	100.0	Q9AHT5_STRPN	Q9Aht5 streptococc
2	799	100.0	Q97RY6_STRPN	Q97ry6 streptococc
3	796	99.6	Q8DQP7_STRPN	Q8dqp7 streptococc
4	795	99.5	Q9S4M8_STRPN	Q9s4m8 streptococc
5	129.5	16.2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	118	14.8	Q8CPK8_STAEP	Q8cpk8 staphylococ
7	112.5	14.1	Q9SP15_PLAFA	Q9sp15 plasmodium
8	112.5	14.1	Q9U6C4_PLAFA	Q9u6c4 plasmodium
9	111.5	14.0	Q25706_PLAFA	Q25706 plasmodium
10	110.5	13.8	Q25705_PLAFA	Q25705 plasmodium
11	110.5	13.8	Q5HQ11_STAEP	Q5hq11 staphylococ
12	110	13.8	Q9U0G0_PLARE	Q9u0g0 plasmodium
13	109.5	13.7	Q5VUJ0_ENTHI	Q5vuj0 entamoeba h
14	109	13.6	Q26019_PLAFA	Q26019 plasmodium
15	109	13.6	Q81436_PLAFA7	Q81436 plasmodium
16	108.5	13.6	Q77355_PLAFA7	Q77355 plasmodium
17	108.5	13.6	Q90784_CHICK	Q90784 gallus gall
18	108.5	13.6	Q512T7_ENTHI	Q512t7 entamoeba h
19	108	13.5	Q25995_PLAFA	Q25995 plasmodium
20	108	13.5	Q81J55_PLAFA7	Q81j55 plasmodium
21	107.5	13.5	Q9FJK9_ARATH	Q9fjk9 arabidopsis
22	107	13.4	Q815F3_PLAFA7	Q815f3 plasmodium
23	106.5	13.3	Q501X8_ENTHO	Q501x8 entamoeba h
24	106.5	13.3	Q5V9M0_PLAXN	Q5v9m0 plasmodium
25	106.5	13.3	Q5W5T1_TETPY	Q5w5t1 tetrahymena
26	106	13.3	Q6BRW2_DEBHA	Q6brw2 debaromyce
27	105	13.1	Q4YMU4_PLABE	Q4ymu4 plasmodium
28	105	13.1	Q96229_PLAFA7	Q96229 plasmodium
29	105	13.1	Q5AGS1_DICDI	Q5ags1 dictyosteli
30	104.5	13.1	Q4IYV2_PLABE	Q4iyv2 plasmodium
31	103.5	13.0	Q44016_DICDI	Q44016 dictyosteli

32	103	12.9	467	2	Q59PE2_CANAL	Q59pe2 candida alb
33	103	12.9	467	2	Q59PL2_CANAL	Q59pl2 candida alb
34	102.5	12.8	674	2	Q7RLI7_PLAYO	Q7rli7 plasmodium
35	102.5	12.8	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
36	102	12.8	540	2	Q54MT2_DICDI	Q54mt2 dictyosteli
37	102	12.8	954	2	Q6HNR0_BACHK	Q6hnr0 bacillus th
38	102	12.8	1603	2	Q5SA14_DICDI	Q5sa14 dictyosteli
39	102	12.8	1603	2	Q8T216_DICDI	Q8t216 dictyosteli
40	101.5	12.7	382	2	Q9V7J0_DROME	Q9v7j0 drosophila
41	101.5	12.7	500	2	Q6BGL7_PASTE	Q6bgl7 paramescium
42	101.5	12.7	556	2	Q9V7I9_DROME	Q9v7i9 drosophila
43	101.5	12.7	785	2	Q9GQ82_DROME	Q9gq82 drosophila
44	101.5	12.7	1042	2	Q7REQ7_PLAYO	Q7req7 plasmodium
45	101	12.6	329	2	Q9NFV9_PLAFA	Q9nfv9 plasmodium

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
EX MEDLINE=21116976; PubMed=11179332;  
RX DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.B., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2S8T.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPTXG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

```
FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 799; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.5e-48;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60
|
|
|
Db 1932 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 1991
|
|
|
QY 61 DAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 120
|
|
|
Db 1992 DAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 2051
|
|
|
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154
|
|
|
Db 2052 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2085
|
|
|

RESULT 2
Q9TRY6_STRPN PRELIMINARY; PRT; 2140 AA.
ID Q9TRY6;
AC Q9TRY6;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAINS=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayal L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007373; AAK74791.1; -; Genomic_DNA.
DR PIR; P95074; P95074.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
```

```
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS0847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; F7A4AD8E2938H334 CRC64;

Query Match 100.0%; Score 799; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-48;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60
|
|
|
Db 1953 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 2012
|
|
|
QY 61 DAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 120
|
|
|
Db 2013 DAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 2072
|
|
|
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154
|
|
|
Db 2073 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2106
|
|
|

RESULT 3
Q8DQF7_STRR6 PRELIMINARY; PRT; 2144 AA.
ID Q8DQF7_STRR6;
AC Q8DQF7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocusNames=spr0551;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Bargett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nickas T.I.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AAK99365.1; -; Genomic_DNA.
DR PIR; A97942; A97942.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
```

DR Pfam; PF02225; PA; 1.  
 DR TIGRfams; TIGR01167; LPTXG anchor; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Cell wall; Signal.  
 FT CHAIN 1 19 Potential.  
 FT SIGNAL 20 2144 Cell wall-associated serine proteinase  
 FT PrtA.  
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 796; DB 2; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 4.2e-48;  
 Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 60  
 |||||  
 Db 1957 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 2016  
 |||||

Qy 61 DAGVYINLSKDTFIKPVFKKIEKKBEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 120  
 |||||  
 Db 2017 DAGVYINLSKDTFIKPVFKKIEKKBEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 2076  
 |||||

Qy 121 EHSQKSDSTKDVTVATVLDKNNISSKSTNNPNK 154  
 |||||  
 Db 2077 EDHSQKSDSTKDVTVATVLDKNNISSKSTNNPNK 2110  
 |||||

RESULT 4  
 ID Q9S4M8 STRPN PRELIMINARY; PRT; 2144 AA.  
 AC Q9S4M8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Cell wall-associated serine proteinase precursor PrtA.  
 GN Name=prtA;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCHI\_TaxID=1313;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3.B;  
 RX MEDLINE=21585565; PubMed=11728722;  
 RA Beche G., Nau R., Wellner A., Hakenbeck R., Reinert R.R., Heinz H.P.,  
 RA Zysk G.;  
 RT "The cell wall-associated serine proteinase PrtA: a highly conserved  
 RT virulence factor of Streptococcus pneumoniae.";  
 RL FEMS Microbiol. Lett. 205:99-104(2001).  
 DR EMBL; AF127143; AAD48399.1; -; Genomic\_DNA.  
 DR HSSP; P00782; 2SER.  
 DR MEROPS; S08.064; -;  
 DR GO; GO:0009986; C:cell surface; IEA.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0042802; F:protein self binding; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR010435; DUP1034.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR InterPro; IPR010259; Prot\_inf\_S8A.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF06280; DUP1034; 1.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.

DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRfams; TIGR01167; LPTXG anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Cell wall; Signal.  
 FT CHAIN 1 19 Potential.  
 FT SIGNAL 20 2144 Cell wall-associated serine proteinase  
 FT PrtA.  
 SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 99.5%; Score 795; DB 2; Length 2144;  
 Best Local Similarity 98.7%; Pred. No. 5e-48;  
 Matches 152; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 60  
 |||||  
 Db 1957 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 2016  
 |||||

Qy 61 DAGVYINLSKDTFIKPVFKKIEKKBEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 120  
 |||||  
 Db 2017 DAGVYINLSKDTFIKPVFKKIEKKBEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 2076  
 |||||

Qy 121 EHSQKSDSTKDVTVATVLDKNNISSKSTNNPNK 154  
 |||||  
 Db 2077 EDHSQKSDSTKDVTVATVLDKNNISSKSTNNPNK 2110  
 |||||

RESULT 5  
 ID Q4XUI6 PLACH PRELIMINARY; PRT; 300 AA.  
 AC Q4XUI6;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN ORFNames=PC000286.03.0;  
 OS Plasmodium chabaudi.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCHI\_TaxID=5825;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,  
 RA Beriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
 RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
 RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
 RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
 RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
 RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
 RT transcriptomic, and proteomic analyses.";  
 RL Science 307:82-86(2005).  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; CAAT01003049; CAH79425.1; -; Genomic\_DNA.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR ProDom; PD000018; WD40; 3.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 4.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Hypothetical protein; Repeat; WD repeat.  
 FT NON TER 1  
 SQ SEQUENCE 300 AA; 34469 MW; 8EDB4E512AFB1945 CRC64;

Query Match 16.2%; Score 129.5; DB 2; Length 300;  
 Best Local Similarity 26.8%; Pred. No. 0.21;  
 Matches 40; Conservative 31; Mismatches 45; Indels 33; Gaps 7;

Qy 9 KPHRVTVTIQNGKMSSTIVSE-----EDFILPVYKGLEKGYQFDGWEISGFEKK 60

Db 130 KVIHTQPSNGKFIASSPDKSIRWSIGDGTLYLAVRGHVGPAYKI-AWSI----- 181  
QY 61 DAGYVINLSKOTFIK-----PVPKKEEKE---PENKPTFVDSKKNDPQVNHSQL 109  
Db 182 DNNYIVSCSDSTLKLWRINHLVPLLLKKEENAEQTKDEK-----NEQKENPQ-NNDP 235  
QY 110 NESHKEDLQREHSOKSDSTKDVATVTL 138  
Db 236 NDEANSEKKKNEKEKNDTKTKTKTL 264

## RESULT 6

Q8CPK8 STAPB PRELIMINARY; PRT; 775 AA.  
AC Q8CPK8  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Penicillin-binding protein 1.  
GN OrderedLocusNames=SE0856;  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=ATCC 12228;  
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;  
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
RA Qin Z.-Q., Miao S.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,  
RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
RT Staphylococcus epidermidis strain (ATCC 12228).";  
RL Mol. Microbiol. 49:1577-1593 (2003).  
DR EMBL; AB016746; AAO04453.1; -; Genomic\_DNA.  
DR HSSP; FL4677; IQME.  
DR GO; GO:0008658; P:penicillin binding; IEA.  
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.  
DR InterPro; IPR005543; PASTA.  
DR InterPro; IPR005311; PBP\_dimer.  
DR InterPro; IPR012338; PBP\_tpept\_fold.  
DR InterPro; IPR001460; Pencil\_bind\_tpept.  
DR Pfam; PF03793; PASTA; 2.  
DR Pfam; PF03717; PBP dimer; 1.  
DR Pfam; PF00905; Transpeptidase; 1.  
DR SMART; SM00740; PASTA; 2.  
KW Complete proteome.  
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.8%; Score 118; DB 2; Length 775;  
Best Local Similarity 27.2%; Pred. No. 3.9;  
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGW--- 51  
Db 615 DSVNAQSLKP-----ITIGNGKQIKQOSVKSQVKGTPKLPKSHKVMWLTGDTMP-DWGTWKE 669  
QY 52 EISGFE-----GKQDAGYVIN--LSKDTFTKPVFKKIEEKEENKPTFDS-----KK 98  
Db 670 DVLAFEDLTPIKIVSTKNGSVFTVNSKISGQIIK-----NKDKIEVLSAEDT 716  
QY 99 KNPQVNHSQLNESHKEDLQREHSOKSDSKDVATVLDKNNISSKSTTN 151  
Db 717 DDQKEDTDESDSNKSKKDKADEDHNTSSSTKN-----DKSNADSKNDSDD 763

## RESULT 7

Q9SP15 PLAFB PRELIMINARY; PRT; 361 AA.  
ID Q9SP15  
AC Q9SP15  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FVO;  
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
RA Hiseaeds H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,  
RA Stowers A.W.,  
RT "Merozoite surface protein 3 and protection against malaria in Aotus  
RT nancymai monkeys.";  
RL J. Infect. Dis. 185:657-664 (2002).  
DR EMBL; AY044180; AAK94780.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAN.  
DR Pfam; PF07133; Merozoite\_SPAN; 1.  
KW Merozoite.  
FT CHAIN <1 >361 merozoite surface protein 3.  
FT NON\_TER 1 1  
FT NON\_TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 14.1%; Score 112.5; DB 2; Length 361;  
Best Local Similarity 22.4%; Pred. No. 4.3;  
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQ----- 47  
Db 147 ETGE-----RNSNNFYTKTKE-----YAGKVKDYERAKNAYKANQAV 187  
QY 48 -----FD---GWEISGP---EGKQDAG-----YVINLSKDTPIKPVFKKIEEKEEN 89  
Db 188 LKAKEASSDYILGWFEFGGVPHEKGEENMLSHLYSVSKOKENISKENDVDVLDK-KEEA 246  
QY 90 KPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSKDVATVLDKQ----- 141  
Db 247 EETEEELKEENKEETESEISEDEHEEEREKEEENDKKQEKQESQNNENNDQKKDME 306  
QY 142 --NISSKSTTN 151  
Db 307 AQNLISKNNNN 318

## RESULT 8

Q9U6C4 PLAFB PRELIMINARY; PRT; 379 AA.  
ID Q9U6C4  
AC Q9U6C4  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymorphic antigen.  
GN Name=ASP-3;  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PCCL/HN;  
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;  
RL Submitted (SFP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF188130; AAF04099.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAN.  
DR Pfam; PF07133; Merozoite\_SPAN; 1.  
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match 14.1%; Score 112.5; DB 2; Length 379;  
Best Local Similarity 23.6%; Pred. No. 4.5;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;  
QY 9 KPHRVTVTIQNGKMSSTIVSEEDF-----ILPVYKGELEKGYQPD-GWEISGP- 56  
Db 171 KPSRLNLSFKRTKSYAEQV--EKDYERAKNAYKANQAVLKAKBASSYDYLGMWFGGV 228

```

Qy 57 -EGKKDAG-----YVILNSKDTPIKPVFKT1EEKKEEKENKPTPDVSKKONPQVNSQLN 110
      |||      |||      |||      |||      |||      |||      |||      |||
Db 229 PEHKGCEENMLSHLYVSSKKONISKENDVDLDE-KEEAEAEETEEEEELEENKSETSEIS 287

Qy 111 -----ESHKREDLQREHSQKSDSTKDTVTATVLDDKKNISKSKTTTN 151

Db 288 EDEEEEEEKEENEKKKEQKEQSNENNDDKKDMEA-----ONLTSKQNNN 336
      |||      |||      |||      |||      |||      |||      |||      |||

```

## RESULT 9

Q25706_PLAFA	
ID	Q25706_PLAFA PRELIMINARY; PRT; 379 AA.
AC	Q25706;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	Polymorphic antigen.
OS	Plasmodium falciparum.
OC	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_Taxid=5833;
RP	[1]
RP	NUCLEOTIDE SEQUENCE.

Query Match	14.0%	Score 111.5;	DB 2;	Length 379;
Best Local Similarity	23.6%;	Pred. No. 5.3;		
Matches 41;	Conservative 35;	Mismatches 59;	Indels 39;	Gaps 8;

Qy	57	-EGKDG----	YVINLSKDT	FIKVP	KIEEK	KEENK	PTFD	VS	KKON	QP	VNHS	QLN	110
		:	:	:	:	:	:	:	:	:	:	:	
Dd	229	PEHKKEEN	MLSH	VSS	KDKEN	ISKEND	DVLDE	-KEE	BAE	ETE	EELE	EEKNE	ETE
		:	:	:	:	:	:	:	:	:	:	:	287

Qy	111	-----ESHKBDLQREKHSQKSDSTKDVTATVLDKNNISSTNN	151
		: :	
Db	288	EDDEEEEEEKEEBENDKKCBQKQBSNNNDQKKDMEA-----QNLISKQNNNN	336

## RESULT 10

Q25705_PLAFA	
ID	Q25705_PLAFA PRELIMINARY; PRT; 379 AA.
AC	Q25705;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	Polymorphic antigen.
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI Taxid:5833;

**Query Match**  
13.8%; Score 110.5; DB 2; Length 379;

```
Best Local Similarity    22.1%, Pred. No. 6.3;
Matches      43; Conservative   35; Mismatches       48; Indels        69; Gaps          9

Qy  1 DTGEVSELKPHRVVTIQNGKMESSTIVSEEDFILPVYKGLELGYQ-----47
     :|||              :|         ||::|||:
Db  167 ETGE-----RNSRNPNFTTKTE---YAGVKDYERAKNAYKANQA V 207
                                           |
Qy  48 -----FD--GWEISGF--EGKKDAG-----YVINLSKDPTFKPKKEEKKBEN 89
     :|||||             :||           |||
Db  208 LKAKEASSYDIWGFCGGVPHEKHGRMLSHLYSSCKENISKENDVDLDE-KEEA 266
                                     |
Qy  90 KPTFDSVEKKONPQNVISQLN-----ESHNKDELORHHSOKSDSTKVDTAT 136
     :| :: |:| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db  267 EETEEEELEEKETTESISEEISEEBBEEBKBEENEKQQOEKSNNENNDAQKMEA- 325
                                   |
Qy  137 VLDRKNISSKSTTNN 151
                       .|.:.|.|
Db  326 ----QNLISKQNNN 336
```

## RESULT 11

Q5HQ11\_STA0Q PRELIMINARY; PRT; 775 AA.

ID Q5HQ11;

AC Q5HQ11;

DT 10-MAY-2005 (TREMBLrel. 30, Created)

DT 10-MAY-2005 (TREMBLrel. 30, Last sequence update)

DT 10-MAY-2005 (TREMBLrel. 30, Last annotation update)

DE Penicillin-binding protein 1.

DE Name=pbpl; OrderedLocNames=SERP0746;

GN Staphylococcus epidermidis (strain ATCC 35984 / RP62A).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=176279;

[1]

NP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RP PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;

RA Gill S.R., Pouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,

RA Ravel J.J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Seaman M.J.,

RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,

RA Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,

RA Dmitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,

RA Hance I.R., Nelson K.E., Fraser C.M.;

RT "Insights on evolution of virulence and resistance from the complete

RT genome analysis of an early methicillin-resistant Staphylococcus

RT aureus strain and a biofilm-producing methicillin-resistant

RT Staphylococcus epidermidis strain.";

RL J. Bacteriol. 187:2426-2438(2005).

DR EMBL; CP000029; AAW54126.1; -; Genomic\_DNA.

DR TIGR; SERP0746; -.

DR GO; GO:0008658; Penicillin binding; IEA.

DR GO; GO:0003273; P:cell wall biosynthesis (sensu Bacteria); IEA.

DR InterPro; IPR005543; PASTA.

DR InterPro; IPR005311; PBP dimer.

DR InterPro; IPR001460; Pencil\_bind\_trypt.

DR Pfam; PF03793; PASTA; 2.

DR Pfam; PF03717; PBP\_dimer; 1.

DR Pfam; PF00505; Transpeptidase; 1.

DR SMART; SM00740; PASTA; 2.

DR Complete proteome.

SW Q5HQ11

Q5HQ11\_STA0Q PRELIMINARY; PRT; 775 AA; 86352 MW; B9395893E0043694 CRC64; SQ

Query Match	13.8%	Score 110.5;	DB 2;	Length 775;
Best Local Similarity	25.0%	Pred. No. 14;		
Matches 42;	Conservative	27;	Mismatches	66;
			Indels	33;
			Gaps	8

53 ISGP--EGKKGAG-----YVINLSKDTFIKPVFKIEEKEEENKPTFDVSKKKNPQVN 105



```

Db 225 FGGVPEHKKENMLSHLVSSKDKENISKENDVDLDE-KEREARETEEELEKKEEET 283
QY 106 HSQLN-----ESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTNN 151
Db 284 ESEISEDEEBEKEEENKKEQKESQSNENDQKDEA-----QNLISKQNNN 337

RESULT 15
Q81436 PLAP7
ID Q81436 PLAP7 PRELIMINARY; PRT; 3008 AA.
AC Q81436;
DT 01-WAR-2003 (TREMBlrel. 23, Created)
DT 01-WAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-WAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein PFE0325w.
GN Name=PFE0325w;

OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sultston J.B.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum."
RL Nature 400:532-538 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.B., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531 (2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAAL5610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0006397; P:mRNA processing; IEA.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 356025 MW; 60BCBBE15C599B4 CRC64;

Query Match 13.6%; Score 109; DB 2; Length 3008;
Best Local Similarity 32.4%; Pred. No. 74;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 64 YVINLSK-----DTFKPVPKKEKKEENKPTFDVSKKQDPQVNHQS----LNSHRE 116
Db 2310 YDIELSKIKFGASIGPFTD-EENKKEENK-EVNKKEENKKEENKKEENKKE 2366

QY 117 DLQREH-----SQKSDSTKDVATVLDKNNISK-----STNNPNK 154
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 16
O77355 PLAP7
ID O77355 PLAP7 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein MAL3P4.20.
GN Name=MAL3P4.20; Synonym=PPC0465c;

```

```

OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sultston J.B.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum."
RL Nature 400:532-538 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.B., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531 (2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAAL5610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0006397; P:mRNA processing; IEA.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 600;
Best Local Similarity 29.3%; Pred. No. 14;
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 27 IVSEEDTLPVY-----KGELEKGVQPDGWEISGFEGKK-----DAGVYVNLSKDTFKPV 77
Db 60 ILGFEDDILYCYISQLKQSEKK---DGEEDKVLNAKGLKINLTFIGNKSDIFIEL 116

QY 78 FKXI--BEKKEE-----ENKPTFDVSK-KQDPQVNHSQLNE-----SHRK 115
Db 117 LELLNEKKEEHIADTLNENK-TNDIKKVNENINENYVNMKDISNKKDKEHVSHQ 175

QY 116 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISKSTNN 150
Db 176 EHNINNNVNLKKEKYTDIQDRKHGRSLQSKSDSYKKRPFPNKRKTSIER-SLSNRYDE 234

QY 151 NPNK 154
Db 235 KTNK 238

RESULT 17
Q80784 CHICK
ID Q80784 CHICK PRELIMINARY; PRT; 1038 AA.
AC Q80784;
DT 01-NOV-1996 (TREMBlrel. 01, Created)

```

```
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Claustrin.
OS Gallus Gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is
RT structurally related to MAP1B.";
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -; mRNA.
DR FIR; JC5497; JC5497.
DR Ensemble; ENSGALG0000014999; Gallus gallus.
SQ SEQUENCE 1038 AA; 117112 MW; 2130694A5B510927 CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 1038;
Best Local Similarity 28.6%; Pred. No. 26;
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 21 KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGEGKQAGYVNLKDTPIKPVK- 79
Db 595 KPETKTIVAKRDV-----TTKEQLGKSTSEKQASEKQDVKPKVKEKVKYKA 646

QY 80 KLEEKKEENKFTDVSKKKNDPQVNSHLSHREKEDLQRE-----EHSQKSDSTKDV 133
Db 647 KPEKKQDEKPKVKEKKEEPLI---KKEPKKEIKVKEKVEKKEKKEKKEV 702

RESULT 18
Q512T7 ENTH1
ID Q512T7 ENTH1 PRELIMINARY; PRT; 1069 AA.
AC Q512T7
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HW-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OC NCBI_TaxID=294381;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harria D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.B., Bhattacharya S., Lohia A.,
RA Foster P.G., Slicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4FA0A8759CB4 CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 27;
Matches 42; Conservative 34; Mismatches 60; Indels 29; Gaps 7;
```

```
QY 4 EVSELKPHRVTVTIQNG-KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGEGKQDA 62
Db 138 EVSKNDVNESSTLTNGEKKLSLSCNEQD-----ELQKSSSSSTD--NKNDKDE 187

QY 63 GYVINLSKDTFIKPVFKKIKBEKKEENK-----PTFDVSKKKQNPQ-----VNHSQLN 110
Db 188 IHFDVVLPKNEKEIKSMEIESSKTEHEKSNLIQPSLNLSEKGNKVESVETAKVLKKSNS 247

QY 111 ESHRKEDLQREHS-QKSDSTKD-----VTATVLDKKNISSKSTT 149
Db 248 NNSGEEDKQDEEVSCERKFDSEKKEKMIKAEVSNKQKVKDKSTT 292

RESULT 19
Q25995 PLAF7
ID Q25995 PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1; -; Genomic DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 354 AA; 40113 MW; 3A7256152F48B527 CRC64;

Query Match 13.5%; Score 108; DB 2; Length 354;
Best Local Similarity 23.2%; Pred. No. 8.8;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFP--EGKQAG-----YVINLS 69
Db 181 VLKAKEASS-----YDYL-----GWEEGGVPEHKKEENMLSHLYVSKD 221

QY 70 KDTPIKPVFKKIKBEKKEE-----ENKPTDVSKKNDPQVNSHLSHREK 116
Db 222 KENTSKENDVDLDKEEAEETEEELKEKNEETSEISEDEEEEEEKEEENDKKK 281

QY 117 DLQREHSQKSDSTKDVATVATVLDKKNISSKSTTNN 151
Db 282 EQEKEQSNENNNDQKDMA-----QLISKQNNN 311

RESULT 20
Q81J55 PLAF7
ID Q81J55 PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN ORFNames=PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,  
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perte M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.G.;  
RA "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum";  
RL Nature 419:498-511(2002).  
DR EMBL; AE014834; AAN35542.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
Query Match 13.5%; Score 108; DB 2; Length 354;  
Best Local Similarity 23.2%; Pred. No. 8.9;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 17 IQNGKSSSTIVSEEDPILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLS 69  
DB 181 VLKAEASS-----YDYL-----GWFGGQVPEHKEENLHLVSSKD 221  
QY 70 KOTFKIPVFKIBKEKBE-----ENKPTFDVSKKONPQVNHSQLNSHRKE 116  
DB 222 KENISKENDVLDKEKEABETEEBLEEKNEBETSEISEDEBEEREEKEENDKK 281  
QY 117 DLQREHSQKSDTKDVTATVLDKNISKSSTNN 151  
DB 282 EQEKEQSNENNQDKOMEA-----QNLISKQNQNN 311  
RESULT 21  
Q9FJK9 ARATH PRELIMINARY; PRT; 470 AA.  
AC Q9FJK9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DB JGI/AN20218.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=99087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RA Tabata S.;  
RA "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned P1 and TAC clones.";  
DR DNA Res. 5:297-308(1998).  
DR EMBL; AB015468; BAB10694.1; -; Genomic DNA.  
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;  
Query Match 13.5%; Score 107.5; DB 2; Length 470;  
Best Local Similarity 20.1%; Pred. No. 13;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
QY 9 KPHRVVTIIONGKEMSTIVSEEDPILPVYKGELEKGYQFDGWEISGFE-----GKK 60  
DB 82 RENRVTDVTQNNNGESK-----YVQDLARRIYDE-EATGQSQAQRIDHPNQK 129

```

NCBI_TaxID=294381;
[1]
NUCLEOTIDE SEQUENCE.
STRAIN=HM-1:IMSS;
Pubmed=15729342; DOI=10.1038/nature03291;
Loftus B., Anderson I., Davies R., Alemark U.C., Samuelson J.,
Amedeo P., Roncaglia P., Berriman M., Hitt R.P., Mann B.J., Nozaki T.,
Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leipzig M.,
Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
Guillen N., Gilchrist C., Sroup S.E., Bhattacharya S., Lohia A.,
Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
Fraser C.M., Hall N.,
The genome of the protist parasite Entamoeba histolytica."
Nature 433:865-869 (2005).
-!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
EMBL: AAPB01001439; EAL42595.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;

Query Match 13.3%; Score 106.5; DB 2; Length 296;
Best Local Similarity 25.3%; Pred. No. 9.3;
Matches 38; Conservative 20; Mismatches 59; Indels 33; Gaps 5;

QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKK 60
Db 142 ESSEGGDSEKHH--DIPTNEGKENDTKDKND-----KSKDNTNEEG-ESSCKEQK 191
QY 61 DAGYINLSKDTFIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQR 120
Db 192 TN-----BESKKTENPQNSSNNKEEQ-----KKEEKKEKKEBQK 228
QY 121 EHSQKSDSTKDVTVLDKNNISSKSTTN 150
Db 229 KEEQKKEEQNQEPINEKNEQENKNTN 258

RESULT 24
QSV9M0 PLAKN PRELIMINARY; PRT; 374 AA.
AC QSV9M0;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DE Merozoite surface protein 5.
GN Name=MSP5;
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H;
RA Black C.G., Wang L., Topolska A.B., Finkelstein D.I., Horne M.K.,
Thomas A.W., Mohandas N., Coppel R.L.;
RT "Merozoite surface proteins 4 and 5 of Plasmodium knowlesi have
differing cellular localisation and association with lipid rafts."
Mol. Biochem. Parasitol. 138:153-158(2004).
RL EMBL: AY573058; RAT77929.1; -; Genomic_DNA.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
SQ SEQUENCE 374 AA; 41642 MW; C0D687C6F23F8989 CRC64;

Query Match 13.3%; Score 106.5; DB 2; Length 374;
Best Local Similarity 25.8%; Pred. No. 12;
Matches 39; Conservative 28; Mismatches 63; Indels 21; Gaps 6;

QY 18 QNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKOAGYVNL-- 68
```

```

Db 19 QSGHPLKFSFWREKKIHLQIYTNLLREEGKNDVNEQMESPSISSTEGKKIQLMISHQL 78
QY 69 -----SKDTPIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQR 123
Db 79 QSGRHQDVSFSLNDSNLKGANEAEAEASGSGKKSDEENYKKSDEENAKKSD---EEN 134
QY 124 SQKSDSTKDVTVLDKNNISSKSS---TTNN 151
Db 135 KDANSNTKDAESAGEENPVSQENQKTLNN 165

RESULT 25
QSV5T1 TETPY PRELIMINARY; PRT; 1015 AA.
AC QSV5T1;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DE Putative P-glycoprotein (Fragment).
GN Name=ndrl;
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymenidae; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GL;
RA Camares O., Denizau F., Bamdad M.;
RT "Characterisation of MDR sequence homologue in Tetrahymena
pyriformis."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AJ514918; CAD55936.2; -; Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transp_like.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS0929; ABC_TM1F; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Nucleotide-binding.
FT NON_TER 1
SQ SEQUENCE 1015 AA; 114219 MW; C82E9AB73768A778 CRC64;

Query Match 13.3%; Score 106.5; DB 2; Length 1015;
Best Local Similarity 27.9%; Pred. No. 35;
Matches 43; Conservative 26; Mismatches 62; Indels 23; Gaps 6;

QY 2 TGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKK 61
Db 308 TGRTTIVIAHLRT-TIRNADE-----ILVIDKGKLVQGTFD--QLIDARGKFE 353
QY 62 AGYINLSKDTFIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQR 120
Db 354 ALAKNQIQKEQKDEERKKKQLQEDRK---DPEERNRKSITVGKSLIMENHLTKQIAA 410
QY 121 EHSQKSDSTKDVTVLDKNNISSKSTTNPNK 154
Db 411 EEEKQKQAYFKE-----LDKNWTRLFTMRPER 439

RESULT 26
```



```
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF001410; RAC71925.2; -; Genomic_DNA.
KW PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8DB89358A84F4F CRC64;

Query Match 13.1%; Score 105; DB 2; Length 951;
Best Local Similarity 24.4%; Pred. No. 42;
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

QY 1 DTGVESELKPHRYT-VTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 EKKINKSLHRLNLSQSGK-----NEQDI-----NKVEKGKQ-----DISNSNAE 171
QY 58 GKGDAGYVNLTKDTPFKPVFKKIEKKE-----EENKPTFD-----VSKKDNPO 103
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 NKCD-----VKEGVKELEKKEEKSIDHKVKNKSDDKHKVENKKSDDHK 219
QY 104 VNHSQNLNHRKEDLQR-EHSHQKSDSTKDVATVLDKNNISSTKTTNNPNK 154
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 VVENKKSDDHKTEEVKVEEHEDEEE-----DKKEKSENKNDENK 262

RESULT 29
Q54GS1 D1C1D1 PRELIMINARY; PRT; 1550 AA.
AC Q54GS1;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=DDB0188660;
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Maderá M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies K., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Degany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulsegged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J.J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostellium discoideum.";
RL Nature 0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF101000221; EAL62484.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1550 AA; 174527 MW; D6D5DF547DC48B6 CRC64;

Query Match 13.1%; Score 105; DB 2; Length 1550;
Best Local Similarity 25.4%; Pred. No. 70;
Matches 46; Conservative 24; Mismatches 59; Indels 52; Gaps 6;
```

```
QY 10 PHRVTVTQNGKEMSTIVSEEDFILPVYKGELEK--GYQFDGWEISGFEGKDGAGYVI- 66
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1328 PHK-----LKKGNRCSSNSSLSILPTLEDHLRKYSPKVDMENTSFDDTLDLKELIS 1383
QY 67 -----NLSKDTFKIPVFKKIEKKEEENKPTFDVS----- 96
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1384 TDELNGNNNNNNEDNNENPFSNIKE NSNEVKVGPVQOEIEKEKEKENIIGEDVEKETS 1443
QY 97 -----KKONPQVNHSQLNHRKEDLQREHSHQKSDSTKDVATVLDKNNISSTKTTN 150
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1444 IEQKEIKKDDNNNNNIVEKEEIEKTDNIKBEIKQDDSNNNKEI-----NN--DNSTTN 1494
QY 151 N 151
Db 1495 N 1495

RESULT 30
Q4YVY2 PLABE PRELIMINARY; PRT; 736 AA.
AC Q4YVY2;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAT1002190; CAH97824.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 736 AA; 85816 MW; A1315CF2D97A6905 CRC64;

Query Match 13.1%; Score 104.5; DB 2; Length 736;
Best Local Similarity 23.0%; Pred. No. 34;
Matches 41; Conservative 34; Mismatches 60; Indels 43; Gaps 8;

QY 17 IONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLTKD----- 72
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401 INNNNDSSS-----GKKNAIPIFLKCKKNNKCDTIETNPWINKAGDLVKKKESLTLL 457
QY 73 FIKPVFKKIEKKEE-----ENKPTFDVSKKDNPOVNH-----SOLNESH----- 114
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
458 FSSPLHKKIEKKQKENVNDYENAKVENIKKNNKGRKIKHPLKIFYFNEENKINTPKPPDD 517
QY 115 ----KEDLQREHSHQKSD-----STKDVATV-----VLDK-----NNISSTKTTNNPNK 154
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
518 IISANENPKTNSPFSKSDILDTSVNNINANYEIEFKGDKIKNSKVDSENIIDNDK 575

RESULT 31
Q44016 D1C1D1 PRELIMINARY; PRT; 325 AA.
AC Q44016;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE G5 ORF.
OS Dictyostellium discoideum (Slime mold).
```







```

Db 251 TEEVPIIE-----NFONYD--NDEDKKEE--VVKSKNE-NQNTTKGMSKPKPKPK 299
QY 92 TFDVSKKKONPOVN-----HSQLNESHKEDLQREHS 124
Db 300 SEASATKVKPOPRKQKKPLSBETVDLTDDLDDDFKDLQLELLEBEEQPKQOI 359
QY 125 QKSDSTKDV-----TATVLDKNNISKSTNNPNK 154
Db 360 KETKSNQSIQKKPSPTIEVDPIAFNDSDEDPDFHTGIKIDCGNNSSSSNNNK 417

RESULT 34
Q7RLE7 PLAYO
ID Q7RLE7_PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RLE7_
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perte M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallon S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome structure and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000712; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001860; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469FE80F7 CRC64;

Query Match 12.8%; Score 102.5; DB 2; Length 674;
Best Local Similarity 23.6%; Pred. No. 43;
Matches 37; Conservative 27; Mismatches 58; Indels 35; Gaps 6;

QY 9 KPHRVTVTQNGKMSSTIVSE-----EDFILPVYKGELEKGYQDGEWISGFEKK 60
Db 490 KVIHTQSPNGKFTASSFDKSIIRWSIGDITYLAVFRGHVPAYKI-AMSI----- 541
QY 61 DAGYVINLSKOTFIK-----PVFKIEE-----KKENKFTFDV--SKKDN 101
Db 542 DNNYIISQSDSTLKLWRINHLVPLLLKKKEENGQPKSDPKNDQKNDQKDKQDKQ 601
QY 102 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVL 138
Db 602 KNDNNDHANNQEDGBEKKKKKKKKKKKKKIKTKILL 638

```

```

RESULT 35
Q9LH98 ARATH
ID Q9LH98_ARATH PRELIMINARY; PRT; 2081 AA.
AC Q9LH98_
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones."
RL DNA Res. 7:217-221(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002057; BAB03174.1; -; Genomic_DNA.
DR HSSP; P01096; 1HF9
DR InterPro; IPR003605; DUF1216.
DR Pfam; PF06746; DUF1216; 2.
SQ SEQUENCE 2081 AA; 232852 MW; D3603E1F85EFPF29 CRC64;

Query Match 12.8%; Score 102.5; DB 2; Length 2081;
Best Local Similarity 27.4%; Pred. No. 1.5e+02;
Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;

QY 14 TTTTQNGKMSSTIVSEDFILPVYKGELEKGYQDGEWISGFEKKD-----AG 63
Db 1659 TVEINGGBELSTERGSKD-----GKIEBGK--EGKENSTKEGSKDKKIEGMEKEN 1708
QY 64 YVINLSKDTFKVPVKIEKKKEENKP--TFDYSK-KDNPOYNHSQLNESHKEDLQ- 119
Db 1709 STKESKDKGKINEIHGDKKEATMEGSKDGGTNGSKDKSKSVKVEINGKDDSLKDSKN 1768
QY 120 ---REHSQKSDSTKDVATVLDKNNISKSTNNPN 153
Db 1769 GDINEINGKEDSVKDNVTIQTGNDNSLTNSSEPN 1805

RESULT 36
Q54MT2 DICDI
ID Q54MT2_DICDI PRELIMINARY; PRT; 540 AA.
AC Q54MT2_
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN OREName=DD0186654;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugeng B., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal R., Kummerfeld S., Madera M., Konfortov B.A., Rivero P.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Parbrother P., Besany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,

```





Search completed: April 24, 2006, 14:59:37  
Job time : 95.5369 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 16.1924 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	799	100.0	2140	2	F95074	serine proteinase,
2	796	99.6	2144	2	A97942	metalloproteinase,
3	110	13.8	558	2	T18467	hypothetical prote
4	108.5	13.6	1038	2	JC5497	claustrin - chicke
5	105	13.1	665	2	B71609	hypothetical prote
6	103.5	13.0	325	2	T18283	hypothetical prote
7	100	12.5	312	2	G81339	probable membrane
8	99	12.4	211	2	T25911	hypothetical prote
9	97.5	12.2	614	2	A84152	hypothetical prote
10	97.5	12.2	1345	2	S46817	hypothetical prote
11	97	12.1	385	2	T20410	hypothetical prote
12	97	12.1	988	2	T14188	hypothetical prote
13	96.5	12.1	540	2	D86432	hypothetical prote
14	96	12.0	535	2	T37189	hypothetical prote
15	96	12.0	2500	2	G71609	hypothetical prote
16	95.5	12.0	644	2	T47835	hypothetical prote
17	95	11.9	348	2	I37271	cyclicin II - human
18	95	11.9	622	2	A90570	lipoprotein (impor
19	95	11.9	1397	2	T10466	DNA topoisomerase
20	94.5	11.8	3724	2	T18427	hypothetical prote
21	94	11.8	210	2	T28771	hypothetical prote
22	93.5	11.7	456	2	T05612	hypothetical prote
23	93	11.6	219	2	B72291	hypothetical prote
24	91	11.4	253	2	T32879	hypothetical prote
25	90.5	11.3	629	2	G96542	hypothetical prote
26	90.5	11.3	670	2	T28391	ORF MSV230 hypothe
27	90.5	11.3	867	2	T27136	hypothetical prote
28	90.5	11.3	871	2	T27135	hypothetical prote
29	90.5	11.3	1332	2	S41552	probable transcrip

30	90.5	11.3	2401	2	T28676	rhoptry protein -
31	90	11.3	645	2	E89883	conserved hypothet
32	90	11.3	1202	1	S05362	probable DNA-direc
33	89.5	11.2	433	2	A89951	trigger factor (im
34	89.5	11.2	508	2	B81594	hypothetical prote
35	89.5	11.2	508	2	E86549	hypothetical prote
36	89.5	11.2	508	2	C72074	hypothetical prote
37	89.5	11.2	700	2	S67610	probable membrane
38	89.5	11.2	1016	2	T19006	ankyrin related pr
39	89	11.1	528	2	E96795	unknown protein P2
40	89	11.1	1888	2	T39009	hypothetical prote
41	88.5	11.1	301	2	T33068	hypothetical prote
42	88.5	11.1	371	2	A71683	hypothetical prote
43	88.5	11.1	443	2	S66040	serine-type D-Ala-
44	88.5	11.1	762	2	G88436	protein T04A8.13 [
45	88.5	11.1	791	2	T24435	hypothetical prote

ALIGNMENTS

RESULT 1  
F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
non, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%	Score	799	DB	2	Length	2140
Best Local Similarity	100.0%	Pred. No.	1.3e-51				
Matches	154	Conservative	0	Mismatches	0	Indels	0
						Gaps	0
Qy	1	DTGEVSELKPHRVTTIQNGKMSSTIVS	EEEDFILPVYKGLKGYQFDGWEISGFEKK	60			
Db	1953	DTGEVSELKPHRVTTIQNGKMSSTIVS	EEEDFILPVYKGLKGYQFDGWEISGFEKK	2012			
Qy	61	DAGVYINLSKDTFKPVFKI	EEKKKEENKPTFDVSKKDNPNVNSHLSHRKEDLQR	120			
Db	2013	DAGVYINLSKDTFKPVFKI	EEKKKEENKPTFDVSKKDNPNVNSHLSHRKEDLQR	2072			
Qy	121	EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK	154				
Db	2073	EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK	2106				

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97942  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2144 <KUR>  
A;Cross-references: UNIPROT:Q8DQ7; UNIPARC:UPI0000E3490; GB:AE007317; PIDN:AAK99365.1;  
C;Genetics:  
A;Gene: pTnA  
C;Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 796; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 2.3e-51;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKVK 60  
Db 1957 DTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKVK 2016

QY 61 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKREDLQR 120  
Db 2017 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKREDLQR 2076

QY 121 EHSQKSDSTKDTVTATVLDKNNISSKSTTNNPK 154  
Db 2077 EDHSQKSDSTKDTVTATVLDKNNISSKSTTNNPK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A;Reference number: Z18937  
A;Accession: T18467  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-558 <LAW>  
A;Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C;Genetics:  
A;Map position: 3  
A;Introns: 84/1; 160/1  
A;Note: C0465C

Query Match 13.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.74;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 27 IVSBEEDFILPVY-----KGELEKGYQDGEWISGFEKVK---DAGYVINLSKDTFIKPV 77  
Db 60 ILGFEDDILYCYISQLKQSKBK--KADGEEDKYLNAXKGLNLTGFIGNKKSDIPBEL 117

QY 78 PKKI--EEKKEE-----ENKPTFDVSK-KKDNPNQVNHSQLNE-----SHRK 115  
Db 118 LELLNEEKKEEHIADTLNENK-TNDIKKVKVENENINENVNENKDISNKKDKEHVSHQN 176

QY 116 E-----DLOREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 150  
Db 177 EHNINNVNLKKEKEYTDIQRDKRHKRSLSQKSDSYKKRPFRNKRKTSTIER-SLSNKRIDE 235

QY 151 NPNK 154  
Db 236 KTNK 239

RESULT 4  
JC5497  
Claustrin - chicken  
N;Alternate names: keratan sulfate proteoglycan  
C;Species: Gallus gallus (chicken)  
C;Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A;Title: Claustarin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A;Reference number: JC5497; MUID:94157526; PMID:7906711  
A;Accession: JC5497  
A;Molecule type: mRNA  
A;Residues: 1-1038 <BURL>  
A;Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:g406318; PID  
A;Accession: PC4334  
A;Molecule type: Protein  
A;Residues: 79-83;299-412;485-502 <BUR2>  
A;Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5  
A;Experimental source: brain  
A;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous  
C;Keywords: chondroitin sulfate proteoglycan, glycoprotein, keratan sulfate  
F;267-270/Region: cell attachment (R-G-D) motif  
F;112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 13.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.9;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 21 KEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKGDAGYVINLSKDTFIKPVFK- 79  
Db 595 KPETKTIVAEKDV-----TTKEEQLGKSETSEKQASEKQDKVPKVKTEKSVKCEVKA 646

QY 80 KIEBKKEENKPTFDVSKKDNPNQVNHSQLNESHKREDLQR-----EHSQKSDSTKDV 133  
Db 647 KPEBKCODEKPKPKKESVKKBEKPLI---KKBEKPKGDIDIKKVKKEVKKBEKKEAKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
A;Accession: B71609  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Portea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: B71609  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-665 <GAR>  
A;Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; NID:AE001362; NID  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0680w

Query Match 13.1%; Score 105; DB 2; Length 665;  
Best Local Similarity 24.4%; Pred. No. 2.1;  
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

QY 1 DTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFE 57  
Db 128 EKNKNSDLHRQNELNLQSK-----NEQDI-----NKNKGRQ-----DISNSNAE 170

QY 58 GKQDAGYVINLSKDTFIKPVFKKIEKKKE-----EENKPTFD-----VSKKDNPNQ 103  
Db 171 NKQD-----VKEGVKLEELKKKEEKISDDHKVNEKKSDDHKVEENKSDDHK 218

QY 104 VNHSQLNESHKREDLQR-EHSQKSDSTKQVTVATVLDKNNISSKSTTNNPK 154  
Db 219 VEENKSDDHKIEEVKVEEHEDEEE-----DKKEKKSENKNKNDENK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C;Species: Dictyostelium discoideum



C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18283  
R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hough  
Genetics 148, 1117-1125, 1998  
A;Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A;Reference number: Z14684; MUID:98198836; PMID:9539429  
A;Accession: T18283  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-325 <RIB>  
A;Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; PI  
C;Genetics:  
A;Introns: 85/1

Query Match 13.0%; Score 103.5; DB 2; Length 325;  
Best Local Similarity 23.8%; Pred. No. 1.2;  
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 11 HRVTVTIONGKEMSTIVSEEDFILPVYK-GEL--EKGYQFDGWEISGFEGK----- 59  
DB 57 HRTITSIKN--RFSVKKIGDEEKLFRISKNGELIVLNELEFDNPHIK--EGKHLRKSQMF 112  
QY 60 ---KDAGYV-----INLSKDTFTKPV-----FKK----- 80  
DB 113 NHIKDSGYATNWEITEIFLESCTLCKBITAQTGRNSYKRNINIKLPBEEEEEEEEEEER 172  
QY 81 --IEEKEEENKPTPDVSKKONQOVNHSQMLNESHKE-----DLQREEHSSQKSDSTK 131  
DB 173 EEEEEQBEVEKPTTISEEEEEETPAVSEEEEEEEEEETPAVSEEEEEKEEQQEEDKEK 232  
QY 132 D 132  
DB 233 D 233

RESULT 7  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81339  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 12.5%; Score 100; DB 2; Length 312;  
Best Local Similarity 25.3%; Pred. No. 2.2;  
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

QY 17 IONGKEMSTI---VSEEDFILPVYK-----GELKGYQFDGWEISGFEGKKGAGYV 65  
DB 24 INQKFLDLDLRDEISSDILRRFRPKKTPNKFELDBEYESKHTKLSNYLYKED---L 80  
QY 66 INL---SKDTFTKPVKLEKEEENKPT---PDVSKKONP-----QVNHSQMLNESHK 115  
DB 81 INVLEEKQSGLAKLIFSKMKERKEENKTKNKFLLFSKCANBEIKNIQTKIQIOTKSNQA 140  
QY 116 EDLQREEHSSQKSDSTKDV--TATVLDKNNISSK--STTNPN 153  
DB 141 TTQTKQEKKELTNSIEKTKQTKETKQKPLIIEKLDVKNQFN 182

RESULT 8

T25911  
hypothetical protein T23B3.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T25911  
R;Maggi, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A;Description: The sequence of C. elegans cosmid T23B3.  
A;Reference number: Z20109  
A;Accession: T25911  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-211 <MAG>  
A;Cross-references: UNIPROT:P91488; UNIPARC:UPI0000007D762; EMBL:U08309; PIDN:AAB42334.1  
A;Experimental source: strain Bristol N2; clone T23B3  
C;Genetics:  
A;Gene: CESP.T23B3.5  
A;Map position: 1  
A;Introns: 30/2; 200/3

Query Match 12.4%; Score 99; DB 2; Length 211;  
Best Local Similarity 29.4%; Pred. No. 1.7;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 57 EGKKGAGYVNLSDTTIKPVFKIKIEKEEENKPTPDVSKKONQOVNH-----SQLNES 112  
DB 69 EGEKKKGKSEKKGDKGKBEKKBEKDGKEDDKKDEKDEKDEKDEKDEKDEKDE 128  
QY 113 HRKEDLQREEHSSQKSDSTKDVATVLDKNNISSKSTTNPNK 154  
DB 129 EKDDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDE 170

RESULT 9  
A84152  
hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C;Accession: A84152  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hir  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: A84152  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-614 <STO>  
A;Cross-references: UNIPROT:Q9K5S1; UNIPARC:UPI000000C4396; GB:AP001520; GB:BA0000004; NI  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH4017

Query Match 12.2%; Score 97.5; DB 2; Length 614;  
Best Local Similarity 27.3%; Pred. No. 7.1;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 6 SELKPHRVTVTIQNGKEMSTIVSE-----EDFILPVYKGELEKGYQFDGW----EISG 55  
DB 386 SEPPE--TYTLQTAIQM-TPIVNEYSPTREEFL-----ARKAQLDQGWADVSKVY 435  
QY 56 PEGKKGAGYVNLSDTTIKPVFKIKIEKK--EENKPTPDVSKK-----DNPQVNHSQL 109  
DB 436 FAGRNIA SQL-----CKIEEEKLQKYNNTVDFPKKEVVNVQSPIKSTAL 482  
QY 110 NESHRKEDLQREEHSSQKSDSTKDVATVLD 139  
DB 483 GKVIGIVDLGRKXHKGEDLERRLSQIE 512

RESULT 10  
S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae

C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C/Accession: S46817  
R:Pavullo, T.  
submitted to the EMBL Data Library, June 1994  
A/Description: The sequence of S. cerevisiae cosmid 9205.  
A/Reference number: S46795  
A/Accession: S46817  
A/Molecule type: DNA  
A/Residues: 1-1345 <FAV>  
A/Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:G500825; PID  
C/Genetics:  
A/Cross-references: SGD:S0001122  
A/Map position: 8R  
C/Superfamily: uncharacterized conserved protein  
C/Keywords: transmembrane protein

Query Match 12.2%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 17;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

Qy 39 KGELEKGYQPDGWEISGPEGK-DAGYVNLKSDTFIKPVFKKIEEKEENKPTFDYSK 97  
Db 1109 KGAIKRG-----SVGEQKVSVDYMLSELRDII-----SRAKSKPKVKYMK 1149

Qy 98 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLD--KNNISSKSTTN---NP 152  
Db 1150 SHDKHRPFHSHVSE-----QKSSSEKSDNDKDLITLTHLDFVQNNFSSSEIFPMWKLSP 1201

Qy 153 NK 154  
Db 1202 QK 1203

RESULT 11  
T20410  
hypothetical protein E02A10.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T20410  
R/Thomas, K.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z19271  
A/Accession: T20410  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-385 <HIL>  
A/Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z81053; PIDN:CAB02877.1;  
A/Experimental source: clone E02A10  
C/Genetics:  
A/Gene: CESP:E02A10.2  
A/Map position: 5  
A/Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.1%; Score 97; DB 2; Length 385;  
Best Local Similarity 29.6%; Pred. No. 4.6;  
Matches 42; Conservative 24; Mismatches 48; Indels 28; Gaps 7;

Qy 1 DTGVESELKPHRVTVTIQ--NGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGPEG 58  
Db 256 ETTDDFCSLOKENVHCTILRIHKEVAEK--NEEDKKEBEPKKEBKEVEKKE---EDE 310

Qy 59 KKDAGVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDL 118  
Db 311 KKDE-----EP--KKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEB 349

Qy 119 QREHSHQKSDSTKDVATVLDK 140  
Db 350 KKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEBKEB 371

RESULT 12  
T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T14188  
R/Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A/Reference number: Z17931  
A/Accession: T14188  
A/Molecule type: DNA  
A/Residues: 1-988 <BEV>  
A/Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI00000A3618; EMBL:AL109819  
A/Experimental source: cultivar Columbia; BAC clone T28D5  
C/Genetics:  
A/Gene: ATSP:T28D5.30  
A/Map position: 4  
A/Introns: 162/3; 201/3; 416/3; 439/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C/Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 12.1%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 13;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

Qy 19 NGKMSSTIVSEDFILPVYKGELEKGY-----QFDGWEISGPEGKDGAVVNLKSDTF 73  
Db 383 NGRQNSNVQSSVDEILSYTDTKVPFSGVGLNVSERDIVELVEDDVRSGAGLSPNVQRDN- 441

Qy 74 IKPVFKKIEEKK-----EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQREE 122  
Db 442 VEPGDDVRSGDMSNPSPAANNVRHGPAFDIMESDNPGRDNVAPMEDHIRSEVQLSP 501

Qy 123 HSQKSDSTKQVT--ATVLDKNNISSKSTTNPNK 154  
Db 502 HVL---GAKDVTDSVSDPTDKVGVDVTDASDPT 532

RESULT 13  
D86432  
hypothetical protein T518.14 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: D86432  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huiztar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: D86432  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-540 <STO>  
A/Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI00000ABF35; GB:AE005172; NID:G4587525; PI  
C/Genetics:  
A/Map position: 1

Query Match 12.1%; Score 96.5; DB 2; Length 540;  
Best Local Similarity 24.8%; Pred. No. 7.3;  
Matches 36; Conservative 25; Mismatches 59; Indels 25; Gaps 5;

Qy 4 EVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGPEGKDGAG 63  
Db 39 BEDESKP-----EGVEKSAFKEESDFADLKESEKK-----ALSDLKSKLEEA 82

Qy 64 YVNLKSDTFIKPVFKK---IEEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120  
Db 83 IVDN-----TLTKTKGSSPKKEKEBKVPEAEVEKKE---EAAEKVEBKESEAVT 136

Qy 121 EEHSHQKSDSTKDVATVLDKNNISS 145

Db 137 BEAPKAEVTVAVVTHEIIPKEVTT 161

## RESULT 14

T37189

hypothetical protein C02H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T37189

C:Description: The sequence of C. elegans cosmid C02H7.

A:Reference number: Z20523

A:Accession: T37189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <LEI>

A:Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:AA47924.1

A:Experimental source: strain Bristol N2; clone C02H7

C:Genetics:

A:Gene: CESP:C02H7.1

A:Map position: X

A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 12.0%; Score 96; DB 2; Length 535;

Best Local Similarity 21.3%; Pred. No. 7.9;

Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

QY 1 DTGEVSELKPHRVTVTIOGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEKCK 60

Db 77 DQSLKNVKAAILI-----SGKDAET-----NKLMLGTWATISFNRN 116

QY 61 DAGYVNLKOTFIKPVFKIEKKEENKPTFVSVKKONPQVNHSQLNSHREKEDLQR 120

Db 117 GTG-----BKKKKVKKEDKGDDEKST---TKRSSKKEHEEKSEKKSAAE 166

QY 121 EHSQKSDSTK-----DVTATVLDKNNISSKSTTNNPNK 154

Db 167 KEKKKSSSSSRHKSRRSEKSEKSKKKEKSTTDEKPK 210

## RESULT 15

G71609

hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004

C:Accession: G71609

C:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.

.; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:9021743; PMID:9804551

A:Accession: G71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2500 <GAR>

A:Cross-references: UNIPROT:Q96223; UNIPARC:UPI0000078301; GB:AE001408; GB:AE001362; NID

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0650w

Query Match

Best Local Similarity 12.0%; Score 96; DB 2; Length 2500;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 9 KPHRVTVTIOGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEKCKAGYVNL 68

Db 2173 KPYKIT---ENNK-----NSGNILKKYSIENEENKKNYDKEQNEICLDKDTQCNYNT 2223

QY 69 -----SKDTFIKPVFKIEKKEENKPTFVSVKKONPQVNHSQLNE---SHRKED 117

Db 2224 KEKNLNDKNSPFSNKKVLEEEKSDDKRD---DKNDNTRKNLNDKNSPFSNKKV 2280

QY 118 LQREHSQKSDSTKQVTVATVLDKNNISSKSTTNNPNK 154

Db 2281 LEBEE---KSDDKRD-----DKNDNTRKNLNDK 2308

## RESULT 16

T47835

hypothetical protein T209.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004

C:Accession: T47835

C:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.

submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24475

A:Accession: T47835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 <NVA>

A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658

A:Experimental source: cultivar Columbia; BAC clone T209

C:Genetics:

A:Map position: 3

A:Introns: 158/2; 329/3

A:Note: T209.90

Query Match 12.0%; Score 95.5; DB 2; Length 644;

Best Local Similarity 23.4%; Pred. No. 11;

Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

QY 18 QNGKEMSSITVSE-----DFILPVYKGELEKGYQFDGWEISGFEKCKAGYVNLKSDTP 73

Db 519 ENSKTEKKTADVADKKSVADFLKRIKNSPQK-----ETTSKNQKNDGNV----- 565

QY 74 IKPVFKIEKKEENKPTFVSVKKONPQVNHSQLNSHREKEDLQREHSQKSDS----- 129

Db 566 -----KENDHQKSKDGNVKKNSKVKPRELSSSTGKKVEVENNNSSKSKKQ 615

QY 130 TKQVTVATVLDKNNISSKSTTNNPNK 154

Db 616 TKETAETVATGKRGESGKDKQPRK 640

## RESULT 17

I37271

C:Species: Homo sapiens (man)

C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004

C:Accession: I37271; S52774

R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

Exp. Cell Res. 216, 174-182, 1995

A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: th

A:Reference number: I37271; MUID:95255491; PMID:7737358

A:Accession: I37271

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-348 <HRS>

A:Cross-references: UNIPROT:Q14093; UNIPARC:UPI0000128C36; EMBL:Z46788; NID:g758586; PI

Query Match 11.9%; Score 95; DB 2; Length 348;

Best Local Similarity 30.9%; Pred. No. 5.8;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 39 KGELEKGYQFDGWEISGFEKCKAGYVNLKSDTFIKPVFKIEKKEENKPTF-----DV 95

Db 205 ESEGEKG---GTEKDSKKGKDS-----KKGKSAIELQAVKADKDEKDKDANKGDE 256

QY 96 SK---KKONPQVNHSQLN-----ESHRKEDLQREHSQKSDSTKD-----VTATVLDKNNI 143

Db 257 SKDAKDAKEIKKGKDKKKPSSTDSKDDVKKE---SKKDATKADKAVAKDTEKESA 313

QY 144 SSK 146

Db 314 DSK 316



T05612  
Hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81F5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.7%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 10;  
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

QY 4 EYSELKP-HRVVTVIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFGKKDA 62  
DB 249 EKDPLKPKIPVSAFLVYANERRAARENKSVVEVAK-----ITGEWKNLSDKKKA 300  
QY 63 GY--VINLSKDTFIKPVFKKIBKKKEEENKPTFDVSKKKDNPNQVNHSQLNESHRKEDLQR 120  
DB 301 PYEKVAKKNKETYLQ-AMEEYRTKEE-----ALSQKE-----EEELLKLHQAELQM 349  
QY 121 EHSQKSDTKDVTATVLDKNNISSKSTNN--PNK 154  
DB 350 LKXGKNTDN-----LIKKEKATKKKKNNVNDPNK 378

RESULT 23  
B72291  
Hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: UNIPROT:Q9XOM6; UNIPARC:UPI00000C12ED; GB:AB001771; GB:AB000512; NID  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1142

Query Match 11.6%; Score 93; DB 2; Length 219;  
Best Local Similarity 26.2%; Pred. No. 4.9;  
Matches 42; Conservative 28; Mismatches 50; Indels 40; Gaps 9;

QY 8 LKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGLKGYQ--FDGWEISG-----PFGK 59  
DB 64 LIPNFVVELYATDEKTLFAKE-----VLGEESVSRYDLFAGFGVGRGTPTPFFPKG 116  
QY 60 KDAGIVIN-LSKDTPIKPVFKKIBKKKEEENKPTFDVSKKKDNPNQVNHSQLNESHRKED- 117  
DB 117 EGLGLVPGVVDKDNFIK-ILKYVAQLKED----FQTYLKKDDPPVGBPLIIEIP-KSDA 170  
QY 118 ---LQREHSQKSDS-----TKDVTATVLDK 140  
DB 171 DFLKEDENAVKVDTPNEVRDRDRIYVTDSPDVATLQEK 210

RESULT 24  
T32879  
Hypothetical protein C17F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T32879  
R:Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C17F3.  
A:Reference number: Z21240  
A:Accession: T32879  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-253 <GAT>  
A:Cross-references: UNIPROT:O44948; UNIPARC:UPI0000074BB9; EMBL:AF043692; PIDN:AAB97531  
A:Experimental source: strain Bristol N2; clone C17F3  
C:Genetics:  
A:Gene: CESP:C17F3.3  
A:Map position: 1  
A:Introns: 41/1

Query Match 11.4%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 8.1;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

QY 76 PVFKKIBKKKE---EENKPTFDVSKK-----KDNQVNHSQLNESHR----- 114  
DB 61 PVAPKVEKKKEKKKEKKADDEKKKTEBKDKKSKKTEBKDKISVKKTKQETKSKDKK 120  
QY 115 ---KEDLQREHSQKSDTKDVTATVLDKNNISSK 146  
DB 121 DEKDEKKESKESKDEKKKDEKDEKKDEKK 155

RESULT 25  
G96542  
Hypothetical protein F17J6.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:Cross-references: UNIPROT:Q9C6P8; UNIPARC:UPI00000A743C; GB:AB005173; NID:g11054631;  
C:Genetics:  
A:Gene: F17J6.14  
A:Map position: 1

Query Match 11.3%; Score 90.5; DB 2; Length 629;  
Best Local Similarity 19.5%; Pred. No. 24;  
Matches 33; Conservative 37; Mismatches 66; Indels 33; Gaps 5;

QY 10 PHRVTVTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFGKKDAGVIVNLS 69  
DB 423 PHLETAKPTKDSANEQTVAEADVAMNPIVEKAMSEWVAEGAANPIVEADG-----A 476  
QY 70 KDTPIKPVFKKIBKKKE---EENKPTFDVSKKKDNPNQ--VNHSQLNESHRKE 116  
DB 477 MNPIVEKAMSQLVEADAAINQAVDANFQTAQTGNDADSDDEPSEPVSHS---ETLNPP 533  
QY 117 DLQREHSQKSDTKDVTATVLDKNN-----ISSKSTTNNPNK 154





A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KEE>  
A:Cross-references: UNIPARC:UPI00000178648; GB:M34281

Query Match 11.3%; Score 90.5; DB 2; Length 2401;  
Best Local Similarity 27.4%; Pred. No. 1.1e+02;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

Qy 16 TIQNGKMSSTI-----VSEBDFILPVYKGELEKGY---QPDGWBISGPEKKDA 62  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 924 TSKNHBEKSIDIRNLSLKIQQFSESYSINDI-KKELEKNVLSEQNNTNDINOYLAKEN 982  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Qy 63 GYVINLSKDTPIKVPFKKIEBKKEEENKPTFDVSKCKONPOVNHSOLNESHRKEDLQREE 122  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 983 IY--NIILNKIKIIIDVKVEYDIEK-----NNKINAEISSEKIITQLKENSSLKE 1035  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Qy 123 HSQKSDSTKO-----VTATVLDKNNISS 145  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 1036 QSKIKSTIDDNVVSECINKITNLKTIVNEKNNT 1072  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 31  
E89883  
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: E89883  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mitutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E89883  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <KUR>  
A:Cross-references: UNIPROT:Q9UUX5; UNIPARC:UPI00000CAAB3; GB:BA000018; PID:g13700929;  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0976

Query Match 11.3%; Score 90; DB 2; Length 645;  
Best Local Similarity 21.4%; Pred. No. 27;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

Qy 4 EVSELKHPRVTV--TIQNGKMSSTIYSEEDFILPVYKGELE-RGYQF-----DGWEIS 54  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 341 KMTDLQDTKYVYVESVENNESMDTFVKH-----PIKTGLNGKKYVMETTTDDYWKDP 395  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Qy 55 GFEGKK-----DAGYVINL-SKDTFI 74  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 396 MVEGQRVRTISKADKANNTRTIFPYVBGKTLDAIVKVHVKTIDYGQYHVRIVDKRAFT 455  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Qy 75 KPVPFKIEEKKEENKPTFDV-----SKCKONPOVNHSOLNESHRKEDLO----REHSQ 125  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 456 KANTDKSNKKEQQDNNAKKGATPATPSKPFTSPVKEKSQKDSOKDONKQLPVSVEKENDA 515  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Qy 126 KSDSTKDTWT-ATVLDKNNISSKSTTNPNPK 154  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 516 SSEGKTKTATKPTKGEVSSSYT--PTK 543  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 32  
S05362  
probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immersus) mitochondrion Ascobolus immersus  
C:Species: mitochondrion Ascobolus immersus  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C:Accession: S05362  
R:Kempken, F.; Meinhardt, F.; Esser, K.





A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: C72074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <ARN>  
A:Cross-references: UNIPROT:Q9Z878; UNIPARC:UPI00000C11C  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: CPn0473

```
Query Match      11.2%; Score 89.5; DB 2; Length 508;  
Best Local Similarity   24.5%; Pred.No. 23;  
Matches          23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;
```

---

Qy	74	KPVPFKLIEKKENKPTFD-----VSKGDNPVNHSQNTHSRK	115
	:	: ::  :::	:
Dd	92	VKGVFYKTTPARPEVSPPRLSPHVHQGLGFGFRDIQRSENPEADLGKMRSVD	151
	:	: ::  :::	:
Qy .	116	EDLOREHSHQSDDSTKDVTATLVKNISSKSTT	149
Dd	152	GDLDRVGHDSDNEDETSRSGEGEPSSKSXS	182
	:	: ::  :::	:

RESULT 37  
S67610  
probable membrane protein YDL074c - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein D2483  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 12-Jul-1996 #sequence revision 12-Jul-1996 #text change 09-Jul-2004

C/Accession: S67610  
R/Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A/Reference number: S67608  
A/Accession: S67610  
A/Molecule type: DNA  
A/Residues: 1-700 <WAM>  
A/Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069EFP; EMBL:274122; NID:g14  
A/Experimental source: strain S288C  
C/Genetics:  
A/Gene: SGD:BRE1; MIPS:YDL074C  
A/Cross-references: SGD:S0002232  
A/Map position: 4L  
C/Keywords: transmembrane protein  
F/69-85/Domain: transmembrane #status predicted <TWM>

RESULT 38  
T19006  
ankyrin related protein C06C3.1 - *Caenorhabditis elegans*  
N;Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T19006; T22086  
R;Berks, M.  
submitted to the EMBL Data Library, August 1994  
A;Reference number: Z19058  
A;Accession: T19006

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1016 <WIL>  
A;Cross-references: UNIPROT:Q17718; UNIPARC:UPI0000009  
A;Experimental source: clone C06C3  
R;Matthews, P.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: Z19510  
A;Accession: T22086  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1016 <WI2>  
A;Cross-references: UNIPARC:UPI0000081D09; EMBL:Z471  
A;Experimental source: clone F42A8  
C;Genetics:  
A;Gene: CHSP:C06C3.1  
A;Map position: 2  
A;Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 1  
C;Keywords: phosphoric monoester hydrolase

	Query Match	11.2%	Score 89.5;	DB 2;	Length 1016;
	Best Local Similarity	21.6%;	Pred. No. 49;		
	Matches	33; Conservative	33; Mismatches	38; Indels	49; Gaps
Qy	15 VTIQNKMSSTIVSEEDFLPVYKGELEKGYPDGMWEISGPGKKDAGVIVNLSDTPI	74			
Db	265 ILLUENGALSD-----LFTGADVLGVADKECIDYLVELA-DTV-	302			
Qy	75 KPVPFKIIEBKKEE---NKPPTFVDSKKONPQNHSQ-LNESHRKBELQ-REESHQ---	126			
Db	303 -----KVQNRKSPGSQPPTSILQEKHRRMPSHEEHVLTSEKRDLQHKQHSENEFL	357			
Qy	127 -----SDSTKDVTAT-VLDKNMISS	145			
Db	358 HSHPSASVGSTSSNTNTTTTIVIGENDISA	390			

RESULT 39  
E96795  
unknown protein F28016.8 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: E96795  
R/Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White-  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.  
ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
C/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: AB6141; MUID:21016719; PMID:11130712

```

Query Match      11.1%; Score 89; DB 2; Length 528;
Best Local Similarity 24.3%; Pred.No. 26;
Matches 35; Conservative 22; Mismatches 51; Indels 36; Gaps 6;

Qy      41  ELEKGQFDGW-----EIGSPKKDAG-----YVNLKDTPIKVPKIER 83
          |||  :  :  :  |||  |||  |||  |||  |||  :  :  :
Db      200  ELAYDNFEWYGANKVRCIGVACSGFLGKRGFQASPYVILLIEDTY---VWSDGDD 256
          |||  :  :  :  |||  |||  |||  |||  |||  :  :  :
Qy      84  KGEENKPTFDVSKGK---DNPQVNHSQLNESHKEDLQREH-----SQKSDSTKD- 132
          |||  :  :  :  :  :  :  :  :  :  :  :  :

```

Db 257 RYSDKIPVYDSELTSEPSKNGESNTNEKEKDISTENHLESTALNIQQSDSTPTP 316  
QY 133 -----VTATVLDKNNISSKSTNN 151  
Db 317 MEEDVVTETVKTETSEDMMKLLSQN 340

RESULT 40  
T39009  
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T39009  
R:Gentles, S.; Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z21815  
A:Accession: T39009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1888 <GEN>  
A:Cross-references: UNIPROT:O14207; UNIPARC:UPI000013AA2P; EMBL:Z98531; PIDN:CAB11064.1;  
A:Experimental source: strain 972h-; cosmid c6B12  
C:Genetics:  
A:Gene: SPDB:SPAC6B12.02c  
A:Map position: 1  
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 11.1%; Score 89; DB 2; Length 1888;  
Best Local Similarity 23.9%; Pred. No. 1.1e+02;  
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

QY 24 SSTIVSEEDP-----ILPVY---KGELEKGYQPDGWEISGFE-----CKKDAG 63  
Db 390 SSSLTSENPFQLNVAANAVSTIPVYRTTKTKMKKN-RFKYVEVEKLPDLILESYGKKAPK 448

QY 64 YVINLSKDTFKPVFKKEEKEENKPTFDVSKKDNPNVNHSQLNESHKEDLQREEH 123  
Db 449 FLRVFARSSSHIP--KMIRRKQMDSKKYFPDKESDRQVIDQVLSDWYSGKHVLQOSH 506

QY 124 SQKSDS-TKDVTATVLDKN-----NISKSTNN 151  
Db 507 SYKPSDSKSGVGNIFSVNSKKHSVINAKTAANN 541

Search completed: April 24, 2006, 15:01:15  
Job time : 18.1924 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 12.0582 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799  
Sequence: 1 DTGVSSELKPHRVTVTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 4: /SIDSS/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 5: /SIDSS/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
  - 8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	14.8	746	6	US-10-793-626-652
2	108	13.5	354	7	US-11-189-817-2
3	106	13.3	651	7	US-11-128-660-1
4	90	11.3	501	6	US-10-485-517-381
5	90	11.3	645	6	US-10-485-517-244
6	90	11.3	886	7	US-11-087-099-11456
7	87	10.9	140	7	US-11-188-298-1015
8	86	10.8	443	7	US-11-188-298-1015
9	86	10.8	700	7	US-11-196-475-74
10	85	10.6	472	6	US-10-793-626-658
11	85	10.6	700	7	US-11-196-475-66
12	85	10.6	708	7	US-11-196-475-66
13	85	10.6	720	6	US-10-793-626-2058
14	84.5	10.6	734	7	US-11-072-512-2272
15	84.5	10.6	1694	7	US-11-052-554A-93
16	83.5	10.5	493	7	US-11-096-568A-3070
17	83.5	10.5	493	7	US-11-096-568A-3071
18	83.5	10.5	510	7	US-11-096-568A-3069
19	82.5	10.3	305	7	US-11-096-568A-17853
20	82.5	10.3	327	7	US-11-096-568A-17852
21	82.5	10.3	386	7	US-11-096-568A-17851
22	82.5	10.3	1036	7	US-11-096-568A-28315
23	82.5	10.3	1070	7	US-11-087-099-5657
24	82.5	10.3	1070	7	US-11-096-568A-28314
25	82.5	10.3	1276	7	US-11-096-568A-28313

26	82	10.3	439	7	US-11-188-298-15964	Sequence 15964, A
27	82	10.3	439	7	US-11-188-298-16606	Sequence 16606, A
28	82	10.3	943	6	US-10-475-204-34	Sequence 34, Appl
29	81.5	10.2	313	7	US-11-096-568A-32043	Sequence 32043, A
30	81.5	10.2	425	7	US-11-096-568A-32042	Sequence 32042, A
31	81.5	10.2	434	7	US-11-096-568A-32041	Sequence 32041, A
32	81.5	10.2	8746	6	US-11-098-686-10232	Sequence 10232, A
33	81	10.1	299	6	US-10-793-626-1888	Sequence 1888, Ap
34	81	10.1	785	6	US-10-793-626-464	Sequence 464, App
35	80.5	10.1	248	6	US-10-793-626-464	Sequence 464, App
36	80	10.0	244	7	US-11-096-568A-11931	Sequence 11931, A
37	79	9.9	339	7	US-11-096-568A-4567	Sequence 4567, Ap
38	79	9.9	568	6	US-10-793-626-2482	Sequence 2482, Ap
39	79	9.9	693	7	US-11-196-475-68	Sequence 68, Appl
40	79	9.9	732	7	US-11-124-368A-248	Sequence 248, Appl
41	79	9.9	803	7	US-11-124-368A-241	Sequence 241, App
42	79	9.9	803	7	US-11-124-368A-242	Sequence 242, App
43	79	9.9	1145	6	US-10-793-626-1432	Sequence 1432, Ap
44	78.5	9.8	258	7	US-11-098-686-10475	Sequence 10475, A
45	78.5	9.8	635	6	US-10-821-234-1573	Sequence 1573, Ap

ALIGNMENTS

RESULT 1  
US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	14.8%	Score 118;	DB 6;	Length 746;
Best Local Similarity	27.2%	Pred. No. 0.026;		
Matches	47;	Conservative	23;	Mismatches 57;
				Indels 46;
				Gaps 9;
Qy	1	DTGVSSELKPHRVTVTIQNGKMSSTIVSEBDFILPVYK-----GELKGYQPDGM--- 51		
Db	586	DSVNAQSLKP-----ITGNGKIQKQSVKSGTKTGPLHPSKVMMLTDTGELTMP-DMTGWTK 640		
Qy	52	RISGFE-----GKDDAGYVIN--LSKDTPTKVPFKIKRKKRKNKPTDVS-----KK 98		
Db	641	DVLAPEDLTKLVKTKNGFVYTNQISIKGQIILK-----NKKLEIVLSAEDT 687		
Qy	99	KDNFQVNHSQLNEHSHRKEDLQREHSHKSGSDTKDTVATVLDKNNISSKSTNN 151		
Db	688	DDQEKTDDESDNKKSKKDAEDHSNTSSSTKN-----DKSNADSKNDSD 734		

RESULT 2  
US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM

```
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
; FILE REFERENCE: 275601US0
; CURRENT APPLICATION NUMBER: US/11/189,817
; CURRENT FILING DATE: 2005-07-27
; PRIOR APPLICATION NUMBER: 60/598,062
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-189-817-2

Query Match      13.5%; Score 108; DB 7; Length 354;
Best Local Similarity 23.2%; Pred. No. 0.073;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQGKEMSTIVSEBDFILPVYKGELEKGYQPDGWEISGF--EGKKDAG-----YVINLS 69
Db 181 VLKAEASS-----DYIL-----GWFGGVPEHKKEENMLSHLYVSSKD 221
QY 70 KDTFKPVFKIEEKKEE-----ENKPTPDVSKKKNPQVNHSQLNESHKRE 116
Db 222 KENISKENDVDLDEKEEBAEETEELEKNEETETSEISEDEEEEBEKEEBENDKK 281
QY 117 DLORREHSQKSTKDVATVTLVDKNISKSSTNN 151
Db 282 EQEKEQSNENNDDQKDMEA-----QNLISKQNNN 311

RESULT 3
US-11-128-660-1
; Sequence 1, Application US/11128660
; Publication No. US20060024324A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmodium
; TITLE OF INVENTION: falciparum
; FILE REFERENCE: 15007dk
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      13.3%; Score 106; DB 7; Length 651;
Best Local Similarity 22.8%; Pred. No. 0.23;
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;

QY 7 ELKPRVTVTIQNGEM-----SSTIVSEBDFILPVYKGELEK----- 44
Db 432 ETEHEETVSGSEPEKADNGVQSNNELNEFV-----ESEKSEHARSKAKEA 485
QY 45 -GYQPD-GWEISGF--EGKKDAG-----YVINLSKDTFKIPVKKIEBKKEENKPTFDV 95
Db 486 SSYDYLGWFGGVPEHKKEENMLSHLYVSSKKNISKENDVDLDE-KEEEAETREE 544
QY 96 SKKKNPQVNHSQLN-----BSHRKEBLQREHSQKSDSTKDVATVTLVDKN 142
Db 545 ELEEKNEETETSEISEDEEBEKEEBEKEEKKGEKEQENNDQKDMEA-----QN 599
QY 143 ISSKSTNN 151
Db 600 LISKQNNN 608

RESULT 4
US-10-485-517-381
```

```
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match      11.3%; Score 90; DB 6; Length 501;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

QY 4 EVSELKPHRVTV--TIQNGKEMSTIVSEBDFILPVYKGELE-KGYQF-----DGWEIS 54
Db 197 KMTDLQDTKYVYVESVENNESMDTFVKH-----PIKTGMLNGKKYMYMVTNDDYWKDF 251
QY 55 GPEGKK-----DAGYVINL-SKDTPI 74
Db 252 MVEGQVRVTTISKDANNRTIIPPYTEGTYDAIVKVVHTIYDGGYHVRIVDKEAPT 311
QY 75 KPVFKKIEEKKEENKPTFDV-----SKKKKNPQVNHSQLNESHKREDLQ-----REEHSQ 125
Db 312 KANTDKSNKKEQDQNSAKKEATPAT?SKPTSPVKESESQKDSQKDDNKLPLSPVEKENDA 371
QY 126 KSDSTKQVT-ATVLDKNISKSSTNNPNK 154
Db 372 SSESQKDKTPATKPTKGEVESSTT--PTK 399

RESULT 5
US-10-485-517-244
; Sequence 244, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 244
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-244

Query Match      11.3%; Score 90; DB 6; Length 645;
Best Local Similarity 21.4%; Pred. No. 5;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;
```



```
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match      10.8%; Score 86; DB 7; Length 700;
Best Local Similarity 23.2%; Pred. No. 12;
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;

QY 13 VVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD-----GWEISGFEKK 60
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKELDKKAIDLQKAQKL 309
QY 61 D-AGYVINLSKDTFKPVFKKIEKKEENKP-TFDVSKKKNPQVN-----HSQIN 110
Db 310 DPAEDNLDIQRDVTREKIQENETNEKKNLPKPGDVSSPKVDKQLQIKESLEDLQEQLK 369
QY 111 ES---HRKEDLQREHSQKSD-----STKQVATVLDKNISSKSTNNPNK 154
Db 370 ETGDNQKREIEKQIEIKSDEKLLKSKDKKASKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 10
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match      10.6%; Score 85; DB 6; Length 472;
Best Local Similarity 32.6%; Pred. No. 9.1;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 81 IBEKKEENKPTFDVSKKKNPQVNHSQLNSHRKEDLQ-----REHSQKSDS--- 129
Db 3 MEENKQPNKE--NWSNKDDNA-----THLNSHRNEDLELFRNKNARQRRRRIDNQSK 56
QY 130 TKQVATAT-----VLDKNISSTNNPNK 154
Db 57 EKDATSTOSQLETKPMDFLDNHKS--HNQNK 86

RESULT 11
US-11-196-475-66
; Sequence 66, Application US/11196475
; Publication No. US20050271682A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Query Match      10.6%; Score 85; DB 7; Length 700;
Best Local Similarity 24.4%; Pred. No. 15;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

QY 13 VVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSK-- 70
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKELDKKA---INLDKQ 306
QY 71 -----DTFKPVFKKIEKKEENKP-TFDVSKKKNPQVN-----HS 107
Db 307 QKLDSADNLDVQRNVTREKIQEDINEKKNLPKPGDVSSPKVDKQLQIKESLEDLQ 366
QY 108 QLNES---HRKEDLQREHSQKSD-----STKQVATVLDKNISSKSTNNPNK 154
Db 367 QLKETGDNQKREIEKQIEIKSDEKLLKSKDKKASKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 12
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 708
```



```
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Query Match
Best Local Similarity 10.6%; Score 85; DB 7; Length 708;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

QY 13 VTTVTQNGKMSSTVSEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVIMLSK-- 70
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
258 ITETIENLRDQLEKATDEE-----HKXIES--QVDAKKCKEKLKKKA---INLDKQAQ 306
QY 71 -----DTPFKPVKKLEKKKEENKP-TDVSCKKONPQVNSQLNESHK 115
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
307 QKLSAENLDVQRDTRVKIQDINEINKNLPKPGDVSSPKVDKQL---QIKES--L 361
QY 116 EDLQRE-----EHSQKSDTKDVTATVLDKNNISK 146
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
362 EDLQRLKAGDENQKRIEIKQIEIKKRDDELLKSK 397

RESULT 13
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match
Best Local Similarity 10.6%; Score 85; DB 6; Length 720;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 81 IEKKKEENKPTFDVSKKD-----NPQVNSQLNESHKEDLQREHSQKSDSTK----D 132
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
38 LEEEQIKALDKKFKASQAKDTNKQNTONNHOKSNKQNSNDKQKQKSNKSKPTKKKEQN 97
QY 133 VTATVLDKNNISSKSTTNPNK 154
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
98 NKGKQONNNKNTKRNKRNK 119

RESULT 14
US-11-072-512-2272
; Sequence 2272, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
```

```
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SERI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2272
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match
Best Local Similarity 10.6%; Score 84.5; DB 7; Length 734;
Matches 39; Conservative 41; Mismatches 58; Indels 41; Gaps 11;

QY 1 DTG-EVSELKPHRVTV-----TIQNGKMSSTIVSE--DFILPVYKG 40
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
169 DTGIVSEVDPSIAAKDGRIGRDLQINGEDVQNRREAVALLSNDECKRIVLLVARP 228
QY 41 ELEKGYQFDGWEISGFGKKDAGVIMLSKDTF-----IKPVFKKIEE-KKEENKPT 92
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
229 EIQLD---EGW-----LEDERNE-FLBELNLEMLEEENEAQPTANEVQPKQKEEBGT 280
QY 93 FDSVSKKONPQVNSQLNESHK-KEDLQREHSQKSDTKDVTATVL-DKNNISKSTT 149
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
281 TDTATSSN---NHEKDSGVGRDTSLENDSSSQENAAEDPNSTLSKSKEDLGQSQDT 336

RESULT 15
US-11-052-554A-83
; Sequence 83, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Haemophilus influenzae Rd
US-11-052-554A-83

Query Match
Best Local Similarity 10.6%; Score 84.5; DB 7; Length 1694;
Matches 24; Conservative 15; Mismatches 15; Indels 3; Gaps 1;

QY 66 INLSKDTFIFKPVFKKIEKKKEENKPTFDVSKKONPQVNSQLNESHKEDLQREHS- 124
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
1288 INTGSATAITETAESKDKPQTETAATEDASQHKANTVADNSVANNSESSDPKSRRSI 1347
QY 125 --QKSDSTKDVATVLDKNNISSKSTTNPNK 154
Db :|||: : :|||: : :|||: : :|||: : :|||: : :|||: : :|||:
```

```
Db 1348 SQPQTSABETTAASTDETTIADNSKRSKPNR 1379

RESULT 16
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

Query Match 10.5%; Score 83.5; DB 7; Length 493;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

QY 16 TQNGKEM--SSTIVSEE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 TQSSGKQTADANTIVTEKKPGKVVPPKKIKTPVSK---KKDETADSNKTETLSKKDDEG 136
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 YVI-----NLKDTFIKPVFKKIEE--KKEENKPTFDVSKKKDNPQVNHs 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 NVAVQAQDDTQSTGKQTANADTTVTPEVKTKGVVPKKQSKTPT---SEKRDN--TADS 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QLNESHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 SKTETKSDKDDKKEER-----VTGEKSGAKTDKLGASDKDVTNVKKG 233
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 17
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3071
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match 10.5%; Score 83.5; DB 7; Length 493;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

QY 16 TQNGKEM--SSTIVSEE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 TQSSGKQTADANTIVTEKKPGKVVPPKKIKTPVSK---KKDETADSNKTETLSKKDDEG 136
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 YVI-----NLKDTFIKPVFKKIEE--KKEENKPTFDVSKKKDNPQVNHs 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 NVAVQAQDDTQSTGKQTANADTTVTPEVKTKGVVPKKQSKTPT---SEKRDN--TADS 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QLNESHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 SKTETKSDKDDKKEER-----VTGEKSGAKTDKLGASDKDVTNVKKG 233
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 18
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match 10.5%; Score 83.5; DB 7; Length 510;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

QY 16 TQNGKEM--SSTIVSEE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 TQSSGKQTADANTIVTEKKPGKVVPPKKIKTPVSK---KKDETADSNKTETLSKKDDEG 153
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 YVI-----NLKDTFIKPVFKKIEE--KKEENKPTFDVSKKKDNPQVNHs 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 NVAVQAQDDTQSTGKQTANADTTVTPEVKTKGVVPKKQSKTPT---SEKRDN--TADS 208
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QLNESHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 SKTETKSDKDDKKEER-----VTGEKSGAKTDKLGASDKDVTNVKKG 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 19
US-11-096-568A-17853
; Sequence 17853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 17853
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(305)
; OTHER INFORMATION: Ceres Seq. ID no. 12361233
US-11-096-568A-17853

Query Match 10.3%; Score 82.5; DB 7; Length 305;
Best Local Similarity 19.6%; Pred. No. 8.7;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;
```

```
Db 137 NVAVQAQDDTQSTGKQTANADTTVTPEVKTKGVVPKKQSKTPT---SEKRDN--TADS 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QLNESHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 SKTETKSDKDDKKEER-----VTGEKSGAKTDKLGASDKDVTNVKKG 233
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 18
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match 10.5%; Score 83.5; DB 7; Length 510;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

QY 16 TQNGKEM--SSTIVSEE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 TQSSGKQTADANTIVTEKKPGKVVPPKKIKTPVSK---KKDETADSNKTETLSKKDDEG 153
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 YVI-----NLKDTFIKPVFKKIEE--KKEENKPTFDVSKKKDNPQVNHs 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 NVAVQAQDDTQSTGKQTANADTTVTPEVKTKGVVPKKQSKTPT---SEKRDN--TADS 208
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QLNESHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 SKTETKSDKDDKKEER-----VTGEKSGAKTDKLGASDKDVTNVKKG 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 19
US-11-096-568A-17853
; Sequence 17853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 17853
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(305)
; OTHER INFORMATION: Ceres Seq. ID no. 12361233
US-11-096-568A-17853

Query Match 10.3%; Score 82.5; DB 7; Length 305;
Best Local Similarity 19.6%; Pred. No. 8.7;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;
```

Qy	10	PHRVTTTQNGKEMSGTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKKDAGVYINUS	69
Db	40	PDLVDCSLNSGDAGSSKKKAEEKSSFRPVAK-----ETPSLEDSNEKKCKTQKAS	87
Qy	70	KDTEFKPVFKLEEKKERENKP-----TFDVSKKKNQPNVHSQLNESHKEDL	118
Db	88	NQHSVK---KDLTEESNESVKEQVGVSTPSYGFSCKDERAEKRREFYSKLEKTHAQEL	144
Qy	119	QREHSQKSDSTKQVATVYLDKN-----TQEEHSSQKSDSTKQVATVYLDKN-----	141
Db	145	EKSNLQAKSKETEBAEKWLKSLNFKATPMPSFYKEPPPPKVELKKIPTTARSFKLGR	204
Qy	142	--NISKSTTNNPN	153
Db	205	SKNTSSGGTEGNPN	218

```

RESULT 20
US-11-096-568A-17852
; Sequence 17852, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17852
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(327)
; OTHER INFORMATION: Ceres Seq. ID no. 12361232
US-11-096-568A-17852

```

```

Query Match      10.3%; Score 82.5; DB 7; Length 327;
Best Local Similarity 19.6%; Pred. No. 9.5;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

Qy    10 PHRVTVTIOCKEMSSITVSEEDFILPVYKGELEKGQFDGWBISFEGKKDAGVYNLS 69
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    62 PDLYDCSLNSGDACSSKKKAESFRPVAK-----TFPSLEDNKKKTOKAS 109
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    70 KDTFIKPVPFKIKBKGBEENKP-----TFDVSKKONQVNHSQLNESHRKEDL 118
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    110 NQHSVK---KOIEBESNESVFPQRVGSTPSYGFSFKCDERAEKRREPFYKLEEKIHAQEL 166
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    119 QREBHSOKSDSTKDVTATVLDKN-----: : : : : : : : : : : : : 141
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    167 EKSNLQAQSKETGEABLKQWLKSLNFKATPMSPSYKGGPPPKVELKKIPTRARSPKLGR 226
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    142 --NISSKSTTNPN 153
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    227 SKNTSSSGTEGNPN 240
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 21
US-11-096-568A-17851
; Sequence 17851, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096, 568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17851

```

```

; LENGTH: 386
; TYPE: PR1
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12361231
US-11-096-568A-17851

```

Query Match	10.3%	Score 82.5;	DB 7;	Length 386;
Best Local Similarity	19.6%;	Pred. No. 12;		
Matches	38;	Conservative 22;	Mismatches 69;	Indels 65; Gaps 4
Qy	10	PHRVTVTIOQKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYINLS	69	
Db	121	PDLVDCSLNGDAGSSKKKAESFRPVAK-----ETPSLEDSNEKKKTQKAS	168	
Qy	70	KDTPIKVPFKKIRBKGBEENKP-----TFDVSKKDKNPNVNHSQLNESHKRED	118	
Db	169	NQHSVK---KDIIEESNESVKPQRVGTPSYGPFKCDERAERKREFYSKLEEKTHAQEL	225	
Qy	119	QRBEHSOKSDSTKDVATVLDKN-----K-----	141	
Db	226	EKSNLQAKSKETEBELKLSLNFKATPMSPFYKEPPPKVELKKIPTTRARSPKLR	285	
Qy	142	--NISKSTTNPN	153	
Db	286	SKNTSSGGTEGNPN	299	

```

RESULT 22
US-11-096-568A-28315
; Sequence 28315, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Pol
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28315
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1036)
; OTHER INFORMATION: Ceres Seq. ID no. 2712010
US-11-096-568A-28315

```

Query Match	10.3%	Score 82.5;	DB 7;	Length 1036;
Best Local Similarity	24.1%	Pred. No. 38;		
Matches	38;	Conservative 29;	Mismatches 68;	Indels 23; Gaps 6
Qy	1	DTGEVS-----ELKPHRVTVITLQNGKMSSTIVSEEDPILPVYKGELEKGYQPDGWEI	53	
Db	504	ETGDTSDPSAKANEQTAKTIVKKIKRVARRKVAEIDNTQ-----DGDSSKKDGSDEKKV	560	
Qy	54	SGPEGKK--DAGYVINLSKDTFIPKVPKCIIEKEEENKPTFDYSSKKDNQVNHSLQNE	111	
Db	561	MEV-GKSSDGSV-----EMPTAESLEDVDENASKTVDYKQETGSPDTKKKEGAS	612	
Qy	112	SHRKBI-LQREH---SOKSDSTKDVTVATVLDKNNISSK	146	
Db	613	SSSKQTKTGEDKKAERKNNSETMSEGGKCIDNNNTDEK	650	

RESULT 23  
US-11-087-099-5657  
; Sequence 5657, Application US/11087099  
; Publication No. US20060041961A1

```
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5657
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Bacillus anthracis str. Ames
US-11-087-099-5657

Query Match          10.3%; Score 82.5; DB 7; Length 1070;
Best Local Similarity 24.1%; Pred. No. 40;
Matches 37; Conservative 24; Mismatches 66; Indels 27; Gaps 6;

QY  2 TGEVSELKP-----HRVTVTIQNGKMSSTIVSEEDFILPVY-----KGELEKGY 46
DB  411 TNEIRDPIQELGKRIKIDAQROKQVFLDEAYMNEEVKIPYDVNGTALQNIWKSEGS 470
QY  47 QPDG---WEISG---FEGKADAGYVINLSKDTFIKPVFKKIEEKKENKPTFDVSKKK- 99
DB  471 ITNGVIKWNSLGERMYEPMWADG-----SKIRFQGRVIGNIVEKREESSNVIQMKLQY 526
QY  100 -DNPQVNSQLNESHRKEDLQREHSQKSDSTKD 132
DB  527 MNKNFERNKVNPTITKSDLLTVKALKITDGKKE 560

RESULT 24
US-11-096-568A-28314
; Sequence 28314, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)-(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

Query Match          10.3%; Score 82.5; DB 7; Length 1070;
Best Local Similarity 24.1%; Pred. No. 40;
Matches 38; Conservative 29; Mismatches 68; Indels 23; Gaps 6;

QY  1 DTGEVS-----ELKPHRVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 53
DB  538 ETGDTSDPSAKANEQTPAKTIKVKKIIKRVAKRVAEIDNKM---DGDSSKKDGSDEKKV 594
QY  54 SGFEGKK--DAGYVINLSKDTFIKPVFKKIEEKKENKPTFDVSKKKDNQVNSQLNE 111
DB  595 MEV-GKKSSDSGSV-----EMKPTAESLEDVKNASKTVDVVKQGTSPDTKKKGAS 646
QY  112 SHRKEDLQREH---SOKSDSTKDVATVLDKNNISSK 146
DB  647 SSSKDKTKTGEDKKAEEKNNSETMSEGGKIDRNNNTDEK 684

RESULT 25
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)-(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313
```

```
Query Match          10.3%; Score 82.5; DB 7; Length 1276;
Best Local Similarity 24.1%; Pred. No. 49;
Matches 38; Conservative 29; Mismatches 68; Indels 23; Gaps 6;

QY  1 DTGEVS-----ELKPHRVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 53
DB  744 ETGDTSDPSAKANEQTPAKTIKVKKIIKRVAKRVAEIDNKM---DGDSSKKDGSDEKKV 800
QY  54 SGFEGKK--DAGYVINLSKDTFIKPVFKKIEEKKENKPTFDVSKKKDNQVNSQLNE 111
DB  801 MEV-GKKSSDSGSV-----EMKPTAESLEDVKNASKTVDVVKQGTSPDTKKKGAS 852
QY  112 SHRKEDLQREH---SOKSDSTKDVATVLDKNNISSK 146
DB  853 SSSKDKTKTGEDKKAEEKNNSETMSEGGKIDRNNNTDEK 890
```

```
RESULT 26
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964
```

```
Query Match          10.3%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 15;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY  45 GYQPDGWEISGFEGKKDAGYVINLSKDTFI-----KP-----VFK 79
DB  45 GISFDGSSVPGFGQIEDSLVFKADPDTYVEVPWNVARVYGYKDNKPYGADPRGILK 104
QY  80 KIEEKKSEEN-----KPTFDVSKKKDN----- 101
DB  105 RALEEKEKGYKAYIGPEPEFYLFKKGNTWELEIPDVGGYFDILTLDKDIRREIAYM 164
QY  102 -----POVNSQLNESHRKEDLQREHSQKSD---STKDVATVLDKNNI 143
DB  165 PSFGLIPEVLHHEVGKQAHEIDFRYDEALKATADNIVSFYKTKVAENHGL 215

RESULT 27
US-11-188-298-16606
```



Qy 55 GFEGKKDAGYVNLKDTF-----IKPVFKKIEBKKEENKPTFDVSKKKNPQVNHQS 108  
Db 338 -----DDPEYAKLAETWGGKKDAEAFDAEKKREESK--DAPAESDAEEAEADD 390  
Qy 109 LNESHRKEDLQREHSQKSDSTKQVAT 136  
Db 391 DNEGDDSDNESKSEETKEAETKEAET 418

## RESULT 31

US-11-096-568A-32041  
; Sequence 32041, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32041  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(434)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592021  
US-11-096-568A-32041

Query Match 10.2%; Score 81.5; DB 7; Length 434;

Best Local Similarity 21.6%; Pred. No. 16;  
Matches 32; Conservative 28; Mismatches 63; Indels 25; Gaps 5;

Qy 7 ELKPHRVTVTIQNGKMSSTIVSEED-----FILPVYK-----GELEKGQYQFGWEIS 54  
Db 287 EWKPKKKNPAVKAKWAKMIDNPFKDDPELYVFPKLYGVVELWQVKSGLFENVLVS 346  
Qy 55 GFEGKKDAGYVNLKDTF-----IKPVFKKIEBKKEENKPTFDVSKKKNPQVNHQS 108  
Db 347 -----DDPEYAKLAETWGGKKDAEAFDAEKKREESK--DAPAESDAEEAEADD 399  
Qy 109 LNESHRKEDLQREHSQKSDSTKQVAT 136  
Db 400 DNEGDDSDNESKSEETKEAETKEAET 427

## RESULT 32

US-11-098-686-10232  
; Sequence 10232, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10232  
; LENGTH: 8746  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10232

Query Match 10.2%; Score 81.5; DB 7; Length 8746;

Best Local Similarity 25.9%; Pred. No. 6.2e+02;  
Matches 38; Conservative 24; Mismatches 60; Indels 25; Gaps 7;  
Qy 22 EMASSTIVSEEDFIL--PVYKGELEKGYQFGWE--ISGPEGKKDAGYVNLKDTFTKPV 77  
Db 3358 EGEDVAEEEDLILLMDTLDNAEEGTKEEHAEVKVGVEGVPDG-----ISEED--KP- 3410  
Qy 78 FKIEEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDL-----QREHSQKS 127  
Db 3411 --KKDKKEEQKATLGDSSGETIERSQQPQREEEKKENSPPSQEESVDE 3468  
Qy 128 DSTKQVATVLDKNNISSKSTTNNPK 154  
Db 3469 TSSVVTSSPLLSINEV--KQTEDEKSAK 3493

## RESULT 33

US-10-793-626-1888  
; Sequence 1888, Application US/10793625  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1888

Query Match 10.1%; Score 81; DB 6; Length 299;

Best Local Similarity 23.4%; Pred. No. 11;  
Matches 37; Conservative 27; Mismatches 60; Indels 34; Gaps 5;

Qy 1 DTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFI----LPVYKGELEKG-YQFGWEISG 55  
Db 35 DT-EKSDKKYHRIISLIPSTIILYRLGIGEDIVGVSTVDDYPKDVKKGKQFDMNLNK 93  
Qy 56 FE-----GKKDAGYVNLKDTFTKPV-----FKKIEEKKEE 87  
Db 94 EELIKAKPDLILAHSQKNSAGKVLKSLKDKGVKVVYVYKDAQSIDETVDTFKSIGQLTDR 153  
Qy 88 ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQ 125  
Db 154 EKQAKELVDETKHNVKXIINSVPKHHKQEVFMEVSSK 191

## RESULT 34

US-10-793-626-264  
; Sequence 264, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 264  
; LENGTH: 785  
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-264

Query Match
Best Local Similarity 10.1%; Score 81; DB 6; Length 785;
Matches 37; Conservative 27; Mismatches 66; Indels 20; Gaps 7;

QY 14 TVTIOGKEMS-STIVSEDFILPVYKGELEKGYQFDGWEISGFGKKGAGY-VINLSKD 71
Db 638 TITERNGNIVCTITVGEEDLMVTNAGVI-----IRLDVHDSQ-NGRAAQGVRLMKUGDG 693
QY 72 TFIKVPFKKIEBKKEEN-----KPTFVSKKKONPQVNHSQLNESHKEDLQREE 122
Db 694 QFVSTVAKVNEEDNEENADEAQSTTTTETAADVEEVD-----DQTGNAIHTEGDAEMES 749
QY 123 -HSQKSDSTKYDTATVLDKNNISSKSTNN 151
Db 750 VEFPENDDRIDIRQDFMDRVNEDIESADN 779

RESULT 35
US-10-793-626-464
; Sequence 464, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 464
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-464

Query Match
Best Local Similarity 10.1%; Score 80.5; DB 6; Length 248;
Matches 40; Conservative 18; Mismatches 63; Indels 41; Gaps 6;

QY 1 DTGEVSELKPHRVTVTIOGKEMS-----STIVSEDFILPVYKGELEKGYQF 48
Db 53 DTNEVHLKEDYLK-TVEN-KKESYDLKGFVDLCNRSIKONEDIL--DYTKLFEARTE 108
QY 49 DGWEISGFGKKGAGYVNLSDTKTFIKVPFKKIEBKKEENKPTFVSKKKONPQVNHQS 108
Db 109 VESDINKAQNKEDA-----SQLSKLENNQQLKDTAKK-----Y 143
QY 109 LNESHRKEDLQREHSQKSDSTKYDTATVLDKNNISSKSTNN 150
Db 144 LNSNNDSDSAKEATKNHISPLIDKQITDINKTNISDNHVDN 185

RESULT 36
US-11-096-568A-11931
; Sequence 11931, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
```

```
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 11931
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(244)
; OTHER INFORMATION: Ceres Seq. ID no. 13659132
US-11-096-568A-11931

Query Match
Best Local Similarity 10.0%; Score 80; DB 7; Length 244;
Matches 40; Conservative 31; Mismatches 56; Indels 42; Gaps 7;

QY 9 KPHRVTVTIOGKEMSSTI-----VSREDFILPVYKGELEKGYQFDGW---EI 53
Db 95 RPVAVASSLRNKKVSSLVDRKKAAGELRDEEBEEDALEYLERKKRKIDGWRKQOI 154
QY 54 SGFEGKKGAGYVNLSDTKTFIKVPFKKIEBKKEENKPTFVSKKKONPQVNHSQLNESH 113
Db 155 ASGEAKENANFV-PLGGDWRDRVKKRAAKKAEKTEPIRAAA-----EQH 199
QY 114 RKG-DLQREHSQKS-----DSTKQVTVLDKNNISSKSTNNPNK 154
Db 200 KGEPLDSELSKGFPSGQAYIDESTKQ-----VYTGNNLTSETTWDPRPSK 244

RESULT 37
US-11-096-568A-4567
; Sequence 4567, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4567
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)..(339)
; OTHER INFORMATION: Ceres Seq. ID no. 15219423
US-11-096-568A-4567

Query Match
Best Local Similarity 9.9%; Score 79; DB 7; Length 339;
Matches 42; Conservative 32; Mismatches 54; Indels 36; Gaps 9;

QY 2 TGEVSELKPHRVTVTIOGKEMSSTIVSE-----EDFILPVYK---GELEKG--Y 46
Db 117 TGDVE--RPRV-YDRRSVGRSNDVKREGGGRGNWGTPTDDIQPVTEPTTEVKS PVA 173
QY 47 QPDGWEISGFGKKGAGYVNL--LSKDTFIKVPFKKIEBKKE-----ENK---PTPD 94
Db 174 EKEGEDATTDAKKEAPEVEQEPEDKEMTLBEYEKILEBKKKALQATKVEERKVDTRK 233
QY 95 VSKKKONPQVNHSQL-----NESHKEDLQREHSQKSDSTKD 132
Db 234 SMQQLSNKKTNDDEEIFIKLGSDKKRKAABEERKAKKSLSLNE 277

RESULT 38
US-10-793-626-2482
; Sequence 2482, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 77.1723 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGVSSELKPHRVTVTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	773	4	US-10-067-385-8
2	799	100.0	2119	3	US-09-769-744A-28
3	799	100.0	2140	4	US-10-282-122A-73670
4	799	100.0	2140	5	US-10-472-928-1180
5	796	99.6	637	5	US-10-617-320-3169
6	615	77.0	117	3	US-09-765-272-68
7	615	77.0	117	6	US-11-106-649-68
8	118	14.8	778	4	US-10-724-972A-5663
9	110.5	13.8	775	4	US-10-282-122A-70721
10	108	13.5	188	5	US-10-691-672A-7
11	107.5	13.5	470	5	US-10-739-930-6262
12	106	13.3	647	5	US-10-691-672A-3
13	105	13.1	665	3	US-09-820-843A-107
14	103.5	13.0	169	5	US-10-691-672A-2
15	103	12.9	707	4	US-10-282-122A-52942
16	101.5	12.7	564	6	US-11-097-143-12723
17	100	12.5	973	5	US-10-732-923-18783
18	97.5	12.2	1184	4	US-10-282-122A-53254
19	97	12.1	3127	5	US-10-732-923-22588
20	96.5	12.1	540	5	US-10-732-923-22820
21	96.5	12.1	2060	4	US-10-381-596A-2
22	95.5	12.0	1373	5	US-10-732-923-16976
23	95	11.9	948	5	US-10-732-923-4286
24	93.5	11.7	898	4	US-10-425-115-205148
25	93	11.6	869	4	US-10-437-963-12282
26	93	11.6	1529	5	US-10-732-923-8762
27	92.5	11.6	903	4	US-10-282-122A-52328

28	92.5	11.6	2468	4	US-10-755-889-615	Sequence 615, App
29	92.5	11.6	2468	5	US-10-489-740-216	Sequence 216, App
30	92.5	11.6	2519	5	US-10-450-763-46995	Sequence 46995, A
31	92	11.5	1005	4	US-10-437-963-187665	Sequence 187665,
32	91.5	11.5	3124	5	US-10-732-923-22709	Sequence 22709, A
33	90.5	11.3	743	4	US-10-171-311-188	Sequence 188, App
34	90.5	11.3	758	3	US-09-925-299-859	Sequence 859, App
35	90.5	11.3	758	3	US-09-925-299-859	Sequence 859, App
36	90.5	11.3	932	4	US-10-282-122A-52510	Sequence 52510, A
37	90	11.3	442	4	US-10-724-972A-5858	Sequence 5858, Ap
38	90	11.3	645	4	US-10-282-122A-70294	Sequence 70294, A
39	90	11.3	645	5	US-10-470-048B-414	Sequence 414, App
40	90	11.3	654	4	US-10-172-502-10	Sequence 10, Appl
41	90	11.3	654	6	US-11-020-509-10	Sequence 3099, Ap
42	90	11.3	1183	6	US-11-097-143-3099	Sequence 3099, Ap
43	89.5	11.2	402	3	US-09-827-664-4	Sequence 4, Appli
44	89.5	11.2	511	4	US-10-289-762-509	Sequence 509, App
45	89.5	11.2	529	3	US-09-827-664-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 799;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 4.1e-61;		
Matches 154;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	DTGVSSELKPHRVTVTIQNGKEMSTIVSEBDFILPVYKGLKGYQFDGWEISGFEGKK	60	
Db	620	DTGVSSELKPHRVTVTIQNGKEMSTIVSEBDFILPVYKGLKGYQFDGWEISGFEGKK	679	
Qy	61	DAGTVINLSKDTFKIPVFKKIEKKKBEKBEKPTFQVSKKONPQVNHSQLNSHREKDLQR	120	
Db	680	DAGTVINLSKDTFKIPVFKKIEKKKBEKBEKPTFQVSKKONPQVNHSQLNSHREKDLQR	739	
Qy	121	BEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	154	
Db	740	BEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	773	

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122W0
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 799; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60
Db      1932 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 1991

QY      61 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db      1992 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2051

QY      121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db      2052 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 799; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60
Db      1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012

QY      61 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db      2013 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2072

QY      121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db      2073 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472923
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 799; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60
Db      1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012

QY      61 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db      2013 DAGYVINLSKDTFIKPVFKKIEEKK3EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2072

QY      121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db      2073 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```





```

; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match 13.5%; Score 108; DB 5; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.17;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

Qy 17 IQNGKMSSTIVSEDFILPVYKGEKGYQPDGWEISGF--EGKKDAG-----YVINLS 69
Db 15 VLKAKEASS-----YDYL-----GWFGGGVPEHKKEENMLSHLYVSSKD 55
Qy 70 KDTPIKVPFKKIEKKEE-----ENKPTPDVSKKKONPQVNHSQLNESHRKE 116
Db 56 KENISKENDVDLDEKEREARETEBELEKKEEETESISEDEBEBEBEKEEBENDKK 115
Qy 117 DLQREHSQKSDSKDVTATVLDKKNISSKSTTNN 151
Db 116 EQEKEQSNENNDDQKQMEA-----QNLISKQNNN 145

RESULT 11
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6262
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270_1.p
US-10-739-930-6262

Query Match 13.5%; Score 107.5; DB 5; Length 470;
Best Local Similarity 20.1%; Pred. No. 0.59;
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

Qy 9 KPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGPF-----GKK 60
Db 82 RERNRVTDTVQNNSGESK-----YVQDLARRIRYDE-EATGSSQAQRIDHPNQK 129
Qy 61 DAGYVINLSKDTPIKVPFKKIEKKEBENKPTPDVSKKKON----- 101
Db 130 NVGITEKAFENSPIEETSHRVDDNNKRNINQKNFTAAKSSSENAKSVSPGADHKRAEVMGK 189
Qy 102 PQVNHSQLNE-----SHRKEDLOREHSQKSDSKDVTATVLDKKNISSKSTTNNPK 154
Db 190 PMENRDQVRQTESAEKSHRKENVTSEKPRQEGVKKTEAKQDKRKEKKEKTESINK 248

RESULT 12
US-10-691-672A-3
; Sequence 3, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRULHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3

```

```
RESULT 14
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication NO. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLRP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match          13.0%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred No. 0.38;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 21 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLSKDTF 73
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 KEASS-----YDYL-----GWFFGGVPEHKKENMLSHLYVSSKDKENI 41

QY 74 IKPVFKHEEKEENKPTFPVSKKDNQVNHSQLN-----ESHKEDLQR 120
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 42 SKENDVDLDE-KEEAEETESEEKNEETESEBSEDEEEEEEKEEENEKKKEQEK 100

QY 121 BEHSQSDSTKDVTVTLVDKNNISSKTTNN 151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 101 EQSNNNDQKQWEA-----QNLISKNNNN 126

RESULT 15
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (359)..(359)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (385)..(385)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (388)..(388)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (396)..(396)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X=any amino acid
```



```
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (404)...(404)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-52942

Query Match
Best Local Similarity 12.9%; Score 103; DB 4; Length 707;
Matches 39; Conservative 25; Mismatches 56; Indels 36; Gaps 5;

QY 1 DTGEVSELPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKK 60
Db 504 DIGDVVEDKD-----TTDKYDS---NKEDIIIPENKSKKAKLFG----- 542
QY 61 DAGYVNLKOTFIKPVFKIPEKKEEN--KPTFDVSKKONPQVNHSQLNESHKEDL 118
Db 543 -----FIKQNEEVEQEEENLNDISPDIIIDKPVENNQVKSEIEIQNELKE-I 589
QY 119 QREHHSQSDSKVDVTATVLDKNNISSKSTTNPNK 154
Db 590 KQEPSQHITEERSVKIEKPIINNLDKVVSSNNESK 625

RESULT 16
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match
Best Local Similarity 12.7%; Score 101.5; DB 6; Length 564;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EDFILPVYKGELEKGYQFDGM-----RISGPEGKKDAGYVI-----NLSKDTPIK 75
Db 78 EDLDTPLSESRFSK--VFDGWDEHRDGHVDQPSGEALDDHDDHDDHDEDEBE 135
QY 76 PVFKKIKREKGEENKPT-----FDVSKKONPQVNHSQLNESHKEDLQREHHSQSDS 129
Db 136 PLTEELEBELEEEPTDEDEPADEEYEEDEENNA--GENITAEADAEDEEEDND 193
QY 130 TKDVTATVLDKNNISSKST 148
```

```
Db 194 EGTVEATVEATTEATTEAT 212
RESULT 17
US-10-732-923-18783
; Sequence 18783, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18783
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(973)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-18783

Query Match
Best Local Similarity 12.5%; Score 100; DB 5; Length 973;
Matches 32; Conservative 39; Mismatches 69; Indels 6; Gaps 3;

QY 7 ELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKKDAGYVI 66
Db 597 KLEYTNINMNDQKGECNSTYKT---LIQHRSSKSGSKGFIHTKINNINREHARGYI 652
QY 67 NLSKDTPIKPVFKIPEKKEENKPTFDVSKKONPQVNHSQLNESHKEDL-QREHHSQ 125
Db 653 SLSKDNNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDN 712
QY 126 KSDSTKQVTATVLDKNNISSKSTTN 151
Db 713 KSEYTFVTATSNKSKODNINKSNSND 737

RESULT 18
US-10-282-122A-53254
; Sequence 53254, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

1 PRIOR APPLICATION NUMBER: 60/230,347  
2 PRIOR FILING DATE: 2000-09-09  
3 PRIOR APPLICATION NUMBER: 60/242,578  
4 PRIOR FILING DATE: 2000-10-23  
5 PRIOR APPLICATION NUMBER: 60/253,625  
6 PRIOR FILING DATE: 2000-11-27  
7 PRIOR APPLICATION NUMBER: 60/257,931  
8 PRIOR FILING DATE: 2000-12-22  
9 PRIOR APPLICATION NUMBER: 60/267,636  
10 PRIOR FILING DATE: 2001-02-09  
11 PRIOR APPLICATION NUMBER: 60/269,308  
12 PRIOR FILING DATE: 2001-02-16  
13 Remaining Prior Application data removed - See File Wrapper or PALM.  
14 NUMBER OF SEQ ID NOS: 78614  
15 SOFTWARE: PatentIn version 3.1  
16 SEQ ID NO 53254  
17 LENGTH: 1184  
18 TYPE: PR1  
19 ORGANISM: Clostridium difficile  
20 US-10-282-122A-53254

Query Match 12.2%; Score 97.5; DB 4; Length 1184;  
Best Local Similarity 26.3%; Pred. No. 14;  
Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
Qy 4 EVSELKPHRVTTIQ-NGKMSSTIV--SEDF--ILPVYKGELEKGYQFDGWEISGPEG 58  
Db 240 EUSEVNEHKVLEKELNEKEQNVKQEDINKEVEVLQDVIEKSVDIYN-SIKGVIS 298  
Qy 59 KIDAGYVNLSDTF-----IKPVFKIEEKE-----ENKPTFDVSKKD 100  
Db 299 KKE--QINLIKERTNFTNRSKLNLEIKDKLNENKQVIELESNK--LSGSELS 354  
Qy 101 NPQVHSQLNESHKEDLOREHSOKSDTKDVTATVLD-KNITSKSTTNPNK 154  
Db 355 TLQENIKVLEGSKQKIKLESANNEILLKESIIDLKQEFNKLSTLNANK 409

RESULT 19  
US-10-732-923-22588  
1 Sequence 22588, Application US/10732923  
2 Publication No. US20050108791A1  
3 GENERAL INFORMATION:  
4 APPLICANT: Edgerton, Michael D  
5 TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
6 FILE REFERENCE: 38-15(52796)C  
7 CURRENT APPLICATION NUMBER: US/10/732,923  
8 PRIOR FILING DATE: 2003-12-10  
9 PRIOR APPLICATION NUMBER: 10/310,154  
10 PRIOR FILING DATE: 2002-12-04  
11 NUMBER OF SEQ ID NOS: 24149  
12 SEQ ID NO 22588  
13 LENGTH: 3127  
14 TYPE: PR1  
15 ORGANISM: Plasmodium falciparum 3D7  
16 US-10-732-923-22588

Query Match 12.1%; Score 97; DB 5; Length 3127;  
Best Local Similarity 24.0%; Pred. No. 50;  
Matches 50; Conservative 32; Mismatches 70; Indels 56; Gaps 8;  
Qy 1 DFGVSELKPHRVTTIQNG-----KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEIS 54  
Db 1974 NTSVLESPLHLIGDIVDNRIKRRKKKKKEIKTIVSDDMPTSPVNIKEYNEQERKKEIV 2033  
Qy 55 G---PEGKKDAGVNLSDTFIKPVFKIEEKEENK-----PTF- 93  
Db 2034 GNLSDYKTKIIPFKTFEGRIK--KKIEKKEKKEKNNNNFLYNDYSSSPKYG 2091  
Qy 94 -----DVSKKKNPQVNHSQL-----NESHK-----EDLOREHSOKS 127  
Db 2092 DNENNFIKIRERKDFQKDFHPNFPNSKFLHNTNPMKNKKNKKNKNNKVRNRYPNYT 2151

Qy 128 DSTKD-VTATVLDKNKNISSKSTTNPNK 154  
Db 2152 SSSKGVSYNFLSDSLFSSDNEYSSUNE 2179

RESULT 20  
US-10-732-923-22820  
1 Sequence 22820, Application US/10732923  
2 Publication No. US20050108791A1  
3 GENERAL INFORMATION:  
4 APPLICANT: Edgerton, Michael D  
5 TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
6 FILE REFERENCE: 38-15(52796)C  
7 CURRENT APPLICATION NUMBER: US/10/732,923  
8 CURRENT FILING DATE: 2003-12-10  
9 PRIOR APPLICATION NUMBER: 10/310,154  
10 PRIOR FILING DATE: 2002-12-04  
11 NUMBER OF SEQ ID NOS: 24149  
12 SEQ ID NO 22820  
13 LENGTH: 540  
14 TYPE: PR1  
15 ORGANISM: Arabidopsis thaliana  
16 US-10-732-923-22820

Query Match 12.1%; Score 96.5; DB 5; Length 540;  
Best Local Similarity 24.8%; Pred. No. 6.4;  
Matches 36; Conservative 25; Mismatches 59; Indels 25; Gaps 5;  
Qy 4 EVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKKDAG 63  
Db 39 EEDESKP-----EGVEKSASFKEESDFADLKESEK-----ALSDLKSLEEA 82  
Qy 64 YVINLSKDTPIKPVFKK---LEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
Db 83 IVDN-----TLTKTKKSSPMKEKKEHVPEAEVEKKCE---EAABEKVEBEKSEAVVT 136  
Qy 121 EEHSQKSDSTKDVATVLDKNKNISS 145  
Db 137 EEAPKAETVEAVVTEIIPKEEVT 161

RESULT 21  
US-10-381-596A-2  
1 Sequence 2, Application US/10381596A  
2 Publication No. US20040014178A1  
3 GENERAL INFORMATION:  
4 APPLICANT: Biostapro AB  
5 TITLE OF INVENTION: von Willebrand factor-binding proteins from  
6 FILE REFERENCE: 110059600  
7 CURRENT APPLICATION NUMBER: US/10/381,596A  
8 CURRENT FILING DATE: 2003-07-02  
9 PRIOR APPLICATION NUMBER: SE 0003573--3  
10 PRIOR FILING DATE: 2000-10-04  
11 NUMBER OF SEQ ID NOS: 17  
12 SOFTWARE: PatentIn Ver. 2.1  
13 SEQ ID NO 2  
14 LENGTH: 2060  
15 TYPE: PR1  
16 ORGANISM: Staphylococcus lugdunensis  
17 US-10-381-596A-2

Query Match 12.1%; Score 96.5; DB 4; Length 2060;  
Best Local Similarity 25.7%; Pred. No. 33;  
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;  
Qy 4 EVSELKPHRVTTIQ-NGKMSSTIVSEEDFILPVYKGELEKGYQFDG--WEISGFEKCK 60  
Db 1914 DIDELEK---ITIVDTNGRE-----IVSRKGQLPP-EQFIGDQWYTGHK--- 1955  
Qy 61 DAGVINLSKDTFTKPVFKIEE-----KKEENKPTFDVSKKKNPQVNHSQLNESH 113  
Db 1956 -----IEKDGIITTYIKKVENAVPAKQLKTKHN--TQSESQFKHTPOVKQQLVKYHN 2006

```

RESULT 25
US-10-437-963-122282
: Sequence 122282, Application US/10437963
: Publication No. US20040123343A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Kovalic, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: APPLICANT: Wu, Wei
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Barbaruk, Brad
: APPLICANT: Li, Ping
: TITLE OF INVENTION: Rice Nucleic Acid Molecules
: TITLE OF INVENTION: Plants and Uses Thereof
: FILE REFERENCE: 38-21(53221)B
: CURRENT APPLICATION NUMBER: US/10/437,963
: CURRENT FILING DATE: 2003-05-14
: NUMBER OF SEQ ID NOS: 204966
: SEQ ID NO 122282
: LENGTH: 869
: TYPE: PRT
: ORGANISM: Oryza sativa
: FEATURE:

```

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_25224C.1.pap  
US-10-437-963-122282

Query Match 11.6%; Score 93; DB 4; Length 869;  
Best Local Similarity 45.1%; Pred. No. 23;  
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY 94 DVSKKKDNPVNHSLNHSRKEQLQREHSQKSDTKDVTATVLDKNN 142  
DB 617 DASKRDNHQSEGNL--SHRDPTKRKKQKTKNATSDACAQEVVTEKNN 665

## RESULT 26

US-10-732-923-8762  
Sequence 8762, Application US/10732923  
Publication No. US20050108791A1  
GENERAL INFORMATION:  
APPLICANT: Edgerton, Michael D  
TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
FILE REFERENCE: 38-15(52796)C  
CURRENT APPLICATION NUMBER: US/10/732,923  
CURRENT FILING DATE: 2003-12-10  
PRIOR APPLICATION NUMBER: 10/310,154  
PRIOR FILING DATE: 2002-12-04  
NUMBER OF SEQ ID NOS: 24149  
SEQ ID NO 8762  
LENGTH: 1529  
TYPE: PRT  
ORGANISM: Plasmodium yoelii yoelii  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(1529)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-732-923-8762

Query Match 11.6%; Score 93; DB 5; Length 1529;  
Best Local Similarity 27.2%; Pred. No. 47;  
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

QY 67 NLKSDT-----PIKPVFKKIEKKEENKPTFDVSKKDNPOVNHSQL-----NES 112  
DB 356 NIISDTQCIKIPKINSEYKKEKKEK-----NEKINDTIHYSESTKNSDNEQ 409

QY 113 H-----RKEDLQREHSQKSDTKDVTATVLDKNN--ISSKSTT 149  
DB 410 HPFLSLRNVKBEKKKKKKTKTKTIVIAQKNKTVIAQKNKT 452

## RESULT 27

US-10-282-122A-52328  
Sequence 52328, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Lianguo  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELTRA 03A4  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 52328  
LENGTH: 903  
TYPE: PRT  
ORGANISM: Clostridium botulinum  
US-10-282-122A-52328

Query Match 11.6%; Score 92.5; DB 4; Length 903;  
Best Local Similarity 25.5%; Pred. No. 27;  
Matches 42; Conservative 27; Mismatches 57; Indels 39; Gaps 9;  
QY 2 TGEVSELKPHR---VTVTIQNGKEMSTVSEBDFILPVYKGELEKGYQPDGWEISGFE 57  
DB 173 SNESIELKNEKSKQLPKINVELKEEKSKQIKKEHNL-----KNEKSKIPKESIKL---K 225  
QY 58 GKGDAGYVINLSKDTFIKPVFKKIEKKEENKPTFDVS-----KKDNPOVNHSQLNE 111  
DB 226 EEKD-----KQVPKEHNSENEENKQISKVNVKLNERSKQLPKIN-VELKE 271  
QY 112 SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN--PNK 154  
DB 272 EKNQSI--KEHNLREBTK----KGLPKVNIELKEBTKKQVPNK 310

## RESULT 28

US-10-755-889-615  
Sequence 615, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469,757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 615  
LENGTH: 2468  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-755-889-615

Query Match 11.6%; Score 92.5; DB 4; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 93;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTFIKPVFKKIEKKEENKPTFDVSKKDNPOVNHSQLNHSRKEQLQRE-----E 122  
DB 638 VKETKVKP-----EDKKEKEKPKKEVAKKEDTKPI---KKEBKPKKEEVKKEVKKBIK 689

Qy 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

## RESULT 29

US-10-489-740-216  
; Sequence 216, Application US/10489740  
; Publication No. US20050112574A1  
; GENERAL INFORMATION:  
; APPLICANT: Bionomics Limited  
; TITLE OF INVENTION: P9  
; FILE REFERENCE: Angiogenesis PCT  
; CURRENT APPLICATION NUMBER: US/10/489,740  
; CURRENT FILING DATE: 2004-03-15  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 216  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-489-740-216

Query Match 11.6%; Score 92.5; DB 5; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 93;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 68 LSKDTFIKPVFKKEEKEENKPTFDVSKKKDPQVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVP-----EDKKKEKPKKEVAKKEDKTPI---KKEKPKKEEVKKEIK 689

Qy 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

## RESULT 30

US-10-450-763-46995  
; Sequence 46995, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:

APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 790C1F3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 46995  
; LENGTH: 2519  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1040)..(1091)  
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1919)..(2122)  
; OTHER INFORMATION: Neuaxin and MAP1B proteins domain identified by PFam,  
; OTHER INFORMATION: accession name MAP1B\_neuraxin, E-values=1.9e-59, PFam score of 190  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(2519)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-46995

Query Match 11.6%; Score 92.5; DB 5; Length 2519;  
Best Local Similarity 31.0%; Pred. No. 95;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 68 LSKDTFIKPVFKKEEKEENKPTFDVSKKKDPQVNHSQLNESHKEDLQRE-----E 122  
Db 689 VKKETKVP-----EDKKKEKPKKEVAKKEDKTPI---KKEKPKKEEVKKEIK 740

Qy 123 HSQKSDSTKDV 133  
Db 741 KEEKKEPKKEV 751

## RESULT 31

US-10-437-963-187665  
; Sequence 187665, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping

FILE REFERENCE: 38-21(53221)B  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 187665  
; LENGTH: 1005  
; TYPE: PRT

ORGANISM: Oryza sativa  
; FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_84346C.1.pep  
US-10-437-963-187665

Query Match 11.5%; Score 92; DB 4; Length 1005;  
Best Local Similarity 26.4%; Pred. No. 34;  
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;

Qy 55 GFEGKDGAGYVINLSKDTFIKPVFKKEEKEENKPTFDVSKK-KONPOV--NHSOLNE 111  
Db 35 GSEKKGSGKNKSIKGTGKQSKELQKESKRSKSTKDKSKKNKDMTQVPTNABPHK 94

Qy 112 SH-----RKE-----DIQREHHSQKSDTKDVTATVLDKNNISSKST 148  
Db 95 EYTTKVKESRTDSSSIQVIGTSSIQEMETNEQVSKSDTSKDMTQVPANAEGIRKEYT 154

Qy 149 T 149  
Db 155 T 155

## RESULT 32

US-10-732-923-22709  
; Sequence 22709, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:

APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10

PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149

SEQ ID NO 22709  
; LENGTH: 3124

```
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-22709

Query Match      11.5%; Score 91.5; DB 5; Length 3124;
Best Local Similarity 23.4%; Pred. No. 1.5e+02;
Matches 48; Conservative 31; Mismatches 73; Indels 53; Gaps 7;

QY 1 DTGEVSELKPHRVTVTIQNG-----KEMSSTIIVSEEDFILPVYKGELEKGYQFDGWEIS 54
Db 1974 NYYLSEPLHLIGDIVDNINIKRKKKCKEIKTIVSDDMFTGVTGKYNKYNEQERKGIIV 2033
QY 55 G---PEGKKDAGVINLSKDTFKVPFKKIEKKEENKPTF----- 93
Db 2034 GNLSDYTKYKICFPFKFKYEGRIKK--NKIEKKEKYNPNFLYNDYSSSPKYGNE 2091
QY 94 -----DVSKCKDNPOVNSQL-----NESHKKE-----DLQREHSQKSDST 130
Db 2092 NNFVIKIRERKDFQKDFHPNFNFSKFLHYNPMKNKKNKNVNVVRNRYPNYTSS 2151
QY 131 KD-VTATVLDKNISKSTNNPNK 154
Db 2152 KDGVSYNFLSDLSFSDNEYSSDNE 2176

RESULT 33
US-10-171-311-188
; Sequence 188, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-188

Query Match      11.3%; Score 90.5; DB 4; Length 743;
Best Local Similarity 23.8%; Pred. No. 32;
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 7 ELKPHRVTVTIQNGKEMSSTIIVSEEDFILPVYKGE-----LEKGYQFDGWEI--SGF 56
Db 281 EARPRR-----QSMKEKEHQVVRNBEHKAQEKGVAQREELVETGQHNVDVEEEAGE 335
QY 57 EGKDKDAGVINLSKDTFKVPFKKIEKKEENKPTFQVSKKDNPNQVNSQLNESHKKE 116
Db 336 EKEKIGIVHSDAE-----KEQEEBQKQEMEVKMEEE-----TEVRESKQQ 378
QY 117 DLQREHS---QKSDSTKDVTA--TVLDKNISKSTNN 151
Db 379 DSQPEVMDVLMVENVKHVIADQEVMTNRVESVPSEN 418
```

```
RESULT 34
US-09-925-299-859
; Sequence 859, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 859
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (590)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-859

Query Match      11.3%; Score 90.5; DB 3; Length 758;
Best Local Similarity 28.1%; Pred. No. 33;
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 3 GEVSELKPHRVTVTIQNGKEMSSTIIVSE--EDFTLPVYKGELEKGYOF-----DGWE--- 52
Db 399 GELOQLSGSL-----HGKSDSPNVYTEKKKIEAIALRLTELERKLTFEQQRSDLWERLY 453
QY 53 -----ISGFEKKDAG-----YVINLSKDTFKVPFKKIEKKEENKPTFQVSKKDD 100
Db 454 VEAKDQNGKQTDGKKKGGRSHRAKNKSETPLGSV-----KETFDAMKNST 501
QY 101 NPQVNH-----SQLNESHKREDLQREHSQKSD--STKQVATVLDK 140
Db 502 KSFVRHHEKIKQAKEA--VKENLKKFSDSVKSTPRHFQKDTTKNIFDE 547

RESULT 35
US-09-925-299-859
; Sequence 859, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 859
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (590)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-859

Query Match      11.3%; Score 90.5; DB 3; Length 758;
Best Local Similarity 28.1%; Pred. No. 33;
```





```

RESULT 39
US-10-470-048B-414
; Sequence 414, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 414
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-414

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 23.9441 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/6 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/H COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/PTUS COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/RE COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	799	100.0	773	2	US-09-590-991-8
2	796	99.6	637	2	US-09-107-433-3169
3	796	99.6	2138	2	US-09-583-110-5274
4	615	77.0	117	2	US-08-961-083-68
5	615	77.0	117	2	US-09-536-784-68
6	615	77.0	117	2	US-09-765-271-68
7	615	77.0	117	2	US-09-765-272A-68
8	118	14.8	746	2	US-09-710-279-652
9	118	14.8	778	2	US-09-134-001C-3868
10	101.5	12.7	347	2	US-09-248-796A-16224
11	95	11.9	348	2	US-09-538-092-1316
12	92.5	11.6	2468	2	US-09-976-594-726
13	92.5	11.6	2468	2	US-09-538-092-1135
14	92.5	11.6	2522	2	US-09-949-016-10237
15	90.5	11.3	743	2	US-08-910-925-3
16	90.5	11.3	743	2	US-09-949-016-6261
17	90.5	11.3	758	2	US-09-949-016-8288
18	90	11.3	442	2	US-09-134-001C-3033
19	90	11.3	654	2	US-10-172-502-10
20	89.5	11.2	402	2	US-09-464-483-4
21	89.5	11.2	402	2	US-09-414-664-4
22	89.5	11.2	511	2	US-09-198-452A-509
23	89.5	11.2	511	2	US-09-438-185A-475
24	89.5	11.2	529	2	US-09-464-483-2
25	89.5	11.2	529	2	US-09-414-664-2
26	89	11.1	280	2	US-09-248-796A-17646
27	87.5	11.0	1702	2	US-08-296-791-5

28	87.5	11.0	1702	2	US-09-839-996-5	Sequence 5, Appli
29	87.5	11.0	1702	2	US-10-080-505-5	Sequence 5, Appli
30	87.5	11.0	1702	2	US-10-645-655-5	Sequence 5, Appli
31	87.5	11.0	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
32	87	10.9	902	2	US-09-134-001C-5157	Sequence 5157, Ap
33	86	10.8	172	2	US-09-248-796A-21065	Sequence 21065, A
34	86	10.8	243	2	US-09-248-796A-20306	Sequence 20306, A
35	86	10.8	262	2	US-09-248-796A-21451	Sequence 21451, A
36	86	10.8	700	2	US-08-235-836C-74	Sequence 74, Appli
37	86	10.8	1315	2	US-09-200-650B-5	Sequence 5, Appli
38	86	10.8	1989	2	US-09-949-016-10076	Sequence 10076, A
39	85.5	10.7	109	2	US-09-248-796A-24668	Sequence 24668, A
40	85.5	10.7	1007	2	US-09-538-092-736	Sequence 736, App
41	85	10.6	465	2	US-09-134-001C-3856	Sequence 3856, Ap
42	85	10.6	472	2	US-09-710-279-658	Sequence 658, App
43	85	10.6	700	1	US-07-720-589-2	Sequence 2, Appli
44	85	10.6	700	1	US-08-785-190-2	Sequence 2, Appli
45	85	10.6	700	2	US-08-235-836C-66	Sequence 66, Appli

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score 799;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 3.8e-75;		
Matches 154;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFEGKK	60	
Db	620	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFEGKK	679	
Qy	61	DAGVINLSKDTFTKPVFKLEEKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQR	120	
Db	680	DAGVINLSKDTFTKPVFKLEEKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQR	739	
Qy	121	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	154	
Db	740	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	773	

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1...637  
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
US-09-107-433-3169  
Query Match 99.6%; Score 796; DB 2; Length 637;  
Best Local Similarity 99.4%; Pred. No. 6.1e-75;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 450 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 509  
QY 61 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR 120  
DB 510 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR 569  
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154  
DB 570 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 603  
RESULT 3  
US-09-583-110-5274  
Sequence 5274, Application US/09583110  
Patent No. 6699703  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al.  
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics  
FILE REFERENCE: PATH00-07A  
CURRENT APPLICATION NUMBER: US/09/583,110  
CURRENT FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/107,433  
PRIOR FILING DATE: 1998-06-30  
PRIOR APPLICATION NUMBER: US 60/085,131  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: US 60/051,553

PRIOR FILING DATE: 1997-07-02  
NUMBER OF SEQ ID NOS: 5322  
SEQ ID NO 5274  
LENGTH: 2138  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5274  
Query Match 99.6%; Score 796; DB 2; Length 2138;  
Best Local Similarity 99.4%; Pred. No. 3.2e-74;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1951 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2010  
QY 61 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR 120  
DB 2011 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR 2070  
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154  
DB 2071 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2104  
RESULT 4  
US-08-961-083-68  
Sequence 68, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961.083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-68  
Query Match 77.0%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.9e-57;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 38 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 97

Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKFTFKPVFKIEEKEENKPTFDVSK 60  
QY 98 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/536,784  
; FILING DATE: 30-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-536-784-68

Query Match 77.0%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.9e-57;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKFTFKPVFKIEEKEENKPTFDVSK 97  
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKFTFKPVFKIEEKEENKPTFDVSK 60  
QY 98 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,271  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/536,784  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-271-68

Query Match 77.0%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.9e-57;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKFTFKPVFKIEEKEENKPTFDVSK 97  
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKFTFKPVFKIEEKEENKPTFDVSK 60  
QY 98 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

## RESULT 7

US-09-765-272A-68  
; Sequence 68, Application US/09765272A  
; Patent No. 6929930  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and  
; NUMBER OF SEQUENCES: 454  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: Dell Latitude C610  
; OPERATING SYSTEM: Windows 2000  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:



Db 260 PFKTKLDQRRB-----WKAKEQANPKKEBENLNQKPVAKQKQKPNSTKKQKQKQK 314  
QY 130 TKDVTATVLDKNNISKSTNNPNK 154  
Db 315 TKKITPKTKSKMLGISTSNLINK 339

RESULT 11  
US-09-538-092-1316  
; Sequence 1316, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurPatSeqFormatter Version 0.9  
; SEQ ID NO 1316  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316

Query Match 11.9%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.096;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
QY 39 KGELEKGQFGWEISGPGKKDAGVINLSKDTPIKPVFKIEEKGEBENKPTF---DV 95  
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KKGDSAIELQNVKADSKKEDGKDKANGDE 256  
QY 96 SK-KKQNPVNHSQLN-----ESHKEDLQREHSQKSDSTKD---VTATVLDKNNI 143  
Db 257 SDAAKDAKBIKKGKDKKKPSSTDSDSKDVKE---SKDATKDAKKVAKKDTKEGSA 313  
QY 144 SSK 146  
Db 314 DSK 316

RESULT 12  
US-09-976-594-726  
; Sequence 726, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 726  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1  
US-09-976-594-726  
Query Match 11.6%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 2.5;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTPIKPVFKIEEKGEBENKPTFDVSKKQNPVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVKP-----EDKKEKEKPKKEVAKKEDKTP1---KKEEKPKKEEVKKEVKKEIK 689  
QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

RESULT 13  
US-09-538-092-1135  
; Sequence 1135, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurPatSeqFormatter Version 0.9  
; SEQ ID NO 1135  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P46821  
US-09-538-092-1135

Query Match 11.6%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 2.5;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTPIKPVFKIEEKGEBENKPTFDVSKKQNPVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVKP-----EDKKEKEKPKKEVAKKEDKTP1---KKEEKPKKEEVKKEVKKEIK 689  
QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

RESULT 14  
US-09-949-016-10237  
; Sequence 10237, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08





```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8288
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8288

Query Match      11.3%; Score 90.5; DB 2; Length 758;
Best Local Similarity 28.1%; Pred. No. 0.82;
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 3 GEVSELKPHRVTVTQNGKEMSTIVSE--EDFILPVYKGELEKGYQP-----DWE--- 52
Db 399 GELOQLSGSQL-----HGKSDSPNVVTEKKK3AILRLRTELEKLTPEQQRSDLWERYL 453
QY 53 -----ISGPEGKKDAG-----YVINLSKDTFIKPVFKKIBKKKEENKPTFDVSKKQD 100
Db 454 VEAKDQNGKQGTGDKKKGGRSHRAKNKSKETFLGVS-----KETFDANKST 501
QY 101 NPQVNH-----SOLNESHKEDLQREHSQKSD--STKDVATVLDK 140
Db 502 KEFVRHHEKIKQAKEA--VKENLKKFSDSVKSTFRHFKDTTKNIFDE 547

RESULT 18
US-09-134-001C-3033
; Sequence 3033, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3033
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3033

Query Match      11.3%; Score 90; DB 2; Length 442;
Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 49; Conservative 29; Mismatches 65; Indels 82; Gaps 11;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44
Db 91 DT-EINPVAQPEVNTQIEKGDFEATVTVPEPVKLDYGLGIEKQETDLSDEELQE 149
QY 45 -----GYQDGEWISG--FEKKDAGVYVINLSKDTFTKP 76
Db 150 SIDHSLSLHAEMVVKEDGAVNGDTVNIIDFSG--SVDGSEFDGQAEYDLGIGSGSFPG 208
QY 77 VPKKIBKKKEENK-----PTFDVS-----KKKDNPOVNHSQLNE--- 111
Db 209 FEEQLEGMTGDEKDVVVTFPEYVHABELAGKEATFKTKVNIKPKQVDFELNDEITANELD 268
QY 112 -----SHRKEDLQREHSQKSDTKDVATVLDKNNISSKSTTN 150
Db 269 SDAENVDEYKLNRLKRLSEKQKATEAENT-----EKEEAINKATEN 308

RESULT 19
US-10-172-502-10
; Sequence 10, Application US/10172502
; Patent No. 6841154
; GENERAL INFORMATION:
; APPLICANT: POSTER, Timothy et al
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.3%; Score 90; DB 2; Length 654;
Best Local Similarity 21.4%; Pred. No. 0.76;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

QY 4 EVSELKPHRVTV--TIQNGKEMSTIVSEEDFILPVYKGELE-KGYQP-----DGWEIS 54
Db 350 KMTDLQDTKYVYVYSEVENNESMDTFVKH-----PIKGTMLNGKKYVMVMTTNDYWKDF 404
QY 55 GPEGKK-----DAGVYINL-SKDTFI 74
Db 405 MVEGQRVRTISKDAKNNTRTIIPVYVEGKTYDAIVKVHVVTIDYDQGHVRIVDKEAPT 464
QY 75 KPVFKKIBKKKEENKPTFDV-----SKKDNPOVNHSQLNESHKEDLQ-----REHSQ 125
Db 465 KANTDKSNKKSQDNSAKKEATPATPSKPTSPVEKESQKQDSQKQDNDKQLPSVEKENDA 524
QY 136 KSDSTQVT-ATVLDKNNISSKSTTNPNK 154
Db 525 SSEGKDKTPATKPTKGEVSSSTT--PTK 552

RESULT 20
US-09-464-483-4
; Sequence 4, Application US/09464483
; Patent No. 6228617
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6228617el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/464,483
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
```

```
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-09-464-483-4

Query Match      11.2%; Score 89.5; DB 2; Length 402;
Best Local Similarity 22.1%; Pred. No. 0.44;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 DETDIKPAQPEVSVTQIEKGKDFIPEATVTVPEVKLGDYKGLIEKQETELSDDELQ 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 -----GYQPDGWEISG--FEGKKDAGVIVNLSKDTTFIKP 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 AIDHSLGHLAEMVVKEDGVVNGDTVNIIDFSG-SVDGEEFEGGQAGYDLBIGSGSFI-P 170
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 77 VFK-----KIEEKK-----EE--NKPTFDVS-----KKKNPQVNHSQLNE- 111
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 GFEEQLEGKMKVDEKDVVTPPEYHAEELAGKEATFKTKVNEIKFKEVPESLTDEIANEL 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 -----SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 DAEANTVDEYKENLRKRLAEQKATDAENV-----EKEEAITKATDN 271
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 22
US-09-198-452A-509
; Sequence 509, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 509
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-509

Query Match      11.2%; Score 89.5; DB 2; Length 511;
Best Local Similarity 24.5%; Pred. No. 0.61;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 74 IKPVFKKIEKKKEENKPTFD-----VSKKDNPNVNHSQLNESHRK 115
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 VKGVFKKTPQARPEVSSPRLPSHVHGQRLPGLGFRDRIQKRSENPEADLGMKRSYSD 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 GDLDRVGHDSNEDSTEDSRS---EGGEPPSSKSS 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 23
US-09-438-185A-475
; Sequence 475, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Versicn 3.0
; SEQ ID NO 475
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
```

```
; FEATURE:
; OTHER INFORMATION: CPn0473
US-09-438-185A-475

Query Match      11.2%; Score 89.5; DB 2; Length 511;
Best Local Similarity 24.5%; Pred. No. 0.64;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 74 IKPVFKKIEKKEENKPTFD-----VSKKNPQVNHSQLNESHK 115
   :|||: : : : : : : : : : : : : : : : : : : : : :
   :|||: : : : : : : : : : : : : : : : : : : : : :
DQ 95 VKGVFKKTPQARPEVSSPLSHVQHQLPGLGFRDRIQKRSNPENADLGKMKRSYSD 154
   :|||: : : : : : : : : : : : : : : : : : : : : :
   :|||: : : : : : : : : : : : : : : : : : : : : :
QY 116 EDLQREHHSOKSDSTKVATVLDKNNISSKSTT 149
   :|||: : : : : : : : : : : : : : : : : : : : : :
   :|||: : : : : : : : : : : : : : : : : : : : : :
DQ 155 GDLDRVGHDSNEDSTEDSR-----EGEPSSKSS 185

RESULT 24
US-09-464-483-2
; Sequence 2, Application US/09464483
; Patent No. 6228617
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6228617el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09464,483
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-464-483-2

Query Match      11.2%; Score 89.5; DB 2; Length 529;
Best Local Similarity 22.1%; Pred. No. 0.64;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 81 DETDIKPAQPEVSVTQIEKGDFIPEATVTVPEVKLGDKYLEIEKQETELSDDELQ 140
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 45 -----GYQFDGWISG--PEGKDGAGVYNLSKDTFIKP 76
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 141 AIDHSLGHLAEMVVKDGVVNGDTWNIDFSG-SVDGEFEGGQAGYDLGSGSFI-P 198
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 77 VPK-----KIEEKK-----NKPTFDS-----KKONPQVNHSQLNE- 111
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 199 GFESQLEGKMKVDEKDVVTVFPFHYHABELAGKATPKTKVNEIKFKEVPBELTDI 258
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 112 -----SHRKEDLQREHHSOKSDSTKVATVLDKNNISSKSTT 150
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 259 DAENTVDYEKENLKRLEAQKATDAENV-----EKEEATKATDN 299
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :

RESULT 26
US-09-414-664-2
; Sequence 2, Application US/09414664
; Patent No. 6242249
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6242249el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,664
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-414-664-2

Query Match      11.2%; Score 89.5; DB 2; Length 529;
Best Local Similarity 22.1%; Pred. No. 0.64;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 81 DETDIKPAQPEVSVTQIEKGDFIPEATVTVPEVKLGDKYLEIEKQETELSDDELQ 140
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 45 -----GYQFDGWISG--PEGKDGAGVYNLSKDTFIKP 76
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 141 AIDHSLGHLAEMVVKDGVVNGDTWNIDFSG-SVDGEFEGGQAGYDLGSGSFI-P 198
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 77 VPK-----KIEEKK-----NKPTFDS-----KKONPQVNHSQLNE- 111
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 199 GFESQLEGKMKVDEKDVVTVFPFHYHABELAGKATPKTKVNEIKFKEVPBELTDI 258
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
QY 112 -----SHRKEDLQREHHSOKSDSTKVATVLDKNNISSKSTT 150
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :
DQ 259 DAENTVDYEKENLKRLEAQKATDAENV-----EKEEATKATDN 299
   : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : :

RESULT 26
```

US-09-248-796A-17646  
; Sequence 17646, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17646  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17646  
Query Match 11.1%; Score 89; DB 2; Length 280;  
Best Local Similarity 25.9%; Pred. No. 0.3;  
Matches 42; Conservative 25; Mismatches 55; Indels 40; Gaps 10;  
QY 3 GEVSELKPHRVTVTQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGPEGKDA 62  
Db 72 GHVSS-----TYTVKRT--SVASTFCSKYDFNVFYSANLSLGF-----ELYSYANKK-- 116  
QY 63 GYVINLSKDTFKIPVKFKIE-EKKEEENKPTFDVSKKONPQVNH-S-QLNESHKEDLQR 120  
Db 117 -----KNSP--PSFEHIEIHSSEENK-----YLGKHPQLQHNLHNLHQRVPIKS 163  
QY 121 EHS-----QKSDSTKDVATVLDKNNISSKSTNNPN 153  
Db 164 HKYEGNRTIINPQNLNDVYHINPTLLSSNG-STSTNNEN 204  
RESULT 27  
US-08-296-791-5  
; Sequence 5, Application US/08296791  
; Patent No. 6245337  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; APPLICANT: Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-839-996-5  
Query Match 11.0%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 5.1;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;  
QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHS- 124  
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355  
QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
US-08-296-791-5  
Query Match 11.0%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 5.1;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;  
QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHS- 124  
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355  
QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154  
Db 1356 SQPQTSABETTAATDDETTIADNSKSKPNR 1387  
RESULT 28  
US-09-839-996-5  
; Sequence 5, Application US/09839996  
; Patent No. 6642371  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; APPLICANT: Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839,996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-839-996-5  
Query Match 11.0%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 5.1;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;  
QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHS- 124  
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355  
QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154











GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 96.6376 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	799	100.0	773	4	AAB48343 S. pneumo
2	799	100.0	2120	3	AAY81710 Streptoco
3	799	100.0	2140	6	ABU01020 S. pneumo
4	799	100.0	2140	6	ABU45746 Protein e
5	799	100.0	2140	8	ADM92113 S pneumon
6	799	100.0	2140	8	ADT50099 S pneumon
7	796	99.6	637	8	ADR94534 Novel S.
8	796	99.6	637	9	AEA58404 Streptoco
9	796	99.6	2138	8	ADK48759 Streptoco
10	615	77.0	117	2	AAW55096 Streptoco
11	615	77.0	117	5	ABF54590 Streptoco
12	615	77.0	117	7	ADC45149 S. pneumo
13	118	14.8	746	5	AAG81779 S. epider
14	118	14.8	778	5	ABP39023 Staphyloc
15	118	14.8	778	6	ADSO6368 Staphyloc
16	110.5	13.8	775	6	ABU42797 Protein e
17	108	13.5	188	9	ADZ79639 P. falcip
18	108	13.5	354	9	ADZ72253 Plasmodi
19	107.5	13.5	470	8	ADT56185 Plant pol
20	107.5	13.5	484	3	AAG47777 Arabidops
21	106	13.3	647	9	ADZ79635 P. falcip
22	106	13.3	651	8	ADOL9012 Amino aci
23	106	13.3	651	8	ADOL9010 P. falcip
24	105	13.1	665	3	AAB18278 Plasmodi

25	105	13.1	665	7	ABO23606 Plasmodi
26	103.5	13.0	169	9	ADZ79634 P. falcip
27	103	12.9	707	6	ABU25018 Protein e
28	101.5	12.7	564	4	ABB61977 Drosophil
29	97.5	12.2	1184	6	ABU25330 Protein e
30	97	12.1	1791	8	ADP25441 Plasmodi
31	96.5	12.1	2060	5	AAE20967 Staphyloc
32	96	12.0	639	9	ADW88472 Staphyloc
33	96	12.0	2500	3	AAB18272 Plasmodi
34	95.5	12.0	635	8	ADW88453 Staphyloc
35	95.5	12.0	635	8	ADW88453 Staphyloc
36	95.5	12.0	643	8	ADW88453 Staphyloc
37	95.5	12.0	643	8	ADW88453 Staphyloc
38	95	11.9	645	9	ADW88453 Staphyloc
39	95	11.9	645	9	ADW88453 Staphyloc
40	95	11.9	645	9	ADW88453 Staphyloc
41	94	11.8	645	9	ADW88453 Staphyloc
42	94	11.8	645	9	ADW88453 Staphyloc
43	94	11.8	645	9	ADW88453 Staphyloc
44	94	11.8	645	9	ADW88453 Staphyloc
45	94	11.8	645	9	ADW88453 Staphyloc

ALIGNMENTS

RESULT 1  
AAB48343  
ID AAB48343 standard; protein; 773 AA.  
XX  
AC AAB48343;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Spi30 polypeptide.  
XX  
KW Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX  
FN WO200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000WO-US015925.  
XX  
PR 10-JUN-1999; 99US-0138453P.  
XX  
(MEDI-) MED IMMUNE INC.  
XX  
PI Adamou JE, Choi GH;  
XX  
XX WPI, 2001-112197/12.  
DR N-PSDB; AAC84742.  
XX  
PT New vaccines comprising Spi28 or Spi30 polypeptides, for treating and  
PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections.  
XX  
PS Claim 8; Page 51-54; 54pp; English.  
XX  
CC The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spl30 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 799; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 6.1e-71;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 620 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 679  
QY 61 DAGVYVNLNLSKDTFIPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 120  
DB 680 DAGVYVNLNLSKDTFIPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 739  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 154  
DB 740 EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 799; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1933 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 1992  
QY 61 DAGVYVNLNLSKDTFIPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 120  
DB 1993 DAGVYVNLNLSKDTFIPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 2052  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 154  
DB 2053 EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Maignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the target sequence  
 CC of the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 799; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTTTQNGKMSSTIVSEEDFLLPVYKGELEKGYQPDGWEISGFEKK 60  
 DB 1953 DTGEVSELKPHRVTTTQNGKMSSTIVSEEDFLLPVYKGELEKGYQPDGWEISGFEKK 2012  
 QY 61 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQR 120  
 DB 2013 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQR 2072  
 QY 121 EEHSQKSDSTKQVTTATVLDKNNISSKSTTNNPNK 154  
 DB 2073 EEHSQKSDSTKQVTTATVLDKNNISSKSTTNNPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.

XX AC ABU45746;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #31273.  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX Streptococcus pneumoniae.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 XX 25-OCT-2001; 2001US-0342923P.  
 XX 08-FEB-2002; 2002US-0072851.  
 XX 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 73670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 799; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTTQNGKMSSTIVSEEDFLLPVYKGELEKGYQPDGWEISGFEKK 60  
 DB 1953 DTGEVSELKPHRVTTTQNGKMSSTIVSEEDFLLPVYKGELEKGYQPDGWEISGFEKK 2012  
 QY 61 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQR 120  
 DB 2013 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQR 2072  
 QY 121 EEHSQKSDSTKQVTTATVLDKNNISSKSTTNNPNK 154  
 DB 2073 EEHSQKSDSTKQVTTATVLDKNNISSKSTTNNPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;  
 XX 03-JUN-2004 (first entry)  
 XX S pneumoniae antigenic protein sequence SeqID310.  
 XX antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 XX antigenic.  
 XX Streptococcus pneumoniae.

PN WO2004020609-A2.  
XX  
PD 11-MAR-2004.  
XX  
PF 02-SEP-2003; 2003WO-US027401.  
XX  
PR 30-AUG-2002; 2002US-0407082P.  
XX  
PA (TUFT ) UNIV TUFTS.  
XX  
PI Camilli A, Hava DL;  
XX  
DR WPI; 2004-239189/22.  
DR N-PSDB; ADW91876.  
XX  
PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
PT diagnosing, treating and preventing active infections of Streptococcus  
PT pneumoniae.  
XX  
PS Claim 27; SEQ ID NO 310; 123pp; English.  
XX  
CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
CC acid molecules and the antigenic polypeptides encoded by them. The  
CC invention may be useful for the production of compounds with an  
CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
CC compositions and methods disclosed are useful for treating Streptococcus  
CC pneumoniae infection. The present sequence is that of an S pneumoniae  
CC protein of the invention.  
XX  
XX Sequence 2140 AA;  
SQ  
Query Match 100.0%; Score 799; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70; Indels 0; Gaps 0;  
Matches 154; Conservative 0; Mismatches 0;  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012  
QY 61 DAGVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
DB 2013 DAGVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2072  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
DB 2073 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106  
RESULT 6  
ADT50099  
ID ADT50099 standard; protein; 2140 AA.  
XX  
AC ADT50099;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
XX  
KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
KW sepsis; meningitis.  
XX  
OS Streptococcus pneumoniae TIGR4.  
XX  
PN WO2004092209-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 15-APR-2004; 2004WO-EP003984.  
XX  
PR 15-APR-2003; 2003EP-00450087.  
XX  
PA (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
PI  
XX  
DR WPI; 2004-758335/74.  
DR N-PSDB; ADT49955.  
XX  
PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
PT treating S. pneumoniae infections.  
XX  
PS Disclosure; SEQ ID NO 177; 191pp; English.  
XX  
CC This invention relates to novel nucleic acids encoding hyperimmune serum  
CC reactive antigens, or fragments derived thereof. Specifically, it refers  
CC to antigens selected from peptides and serum reactive epitopes that can  
CC be used in pharmaceutical compositions that exhibit antibacterial  
CC activity. The present invention describes a composition (including the  
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
CC that is useful for manufacturing a medicament such as a vaccine, which  
CC can be used to treat or prevent bacterial infections, particularly S.  
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
CC be used for isolating, purifying and/ or identifying an interaction  
CC partner of the hyperimmune serum reactive antigen, as well as for  
CC manufacturing a functional nucleic acid selected from aptamers and  
CC Spiegelmers or for manufacturing a functional ribonucleic acid selected  
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
CC of the invention.  
XX  
XX Sequence 2140 AA;  
SQ  
Query Match 100.0%; Score 799; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70; Indels 0; Gaps 0;  
Matches 154; Conservative 0; Mismatches 0;  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012  
QY 61 DAGVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
DB 2013 DAGVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2072  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
DB 2073 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106  
RESULT 7  
ADR94534  
ID ADR94534 standard; protein; 637 AA.  
XX  
AC ADR94534;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
XX  
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
KW bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US6800744-B1.  
XX  
PD 05-OCT-2004.  
XX  
PF 30-JUN-1998; 98US-00107433.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
XX  
PA 12-MAY-1998; 98US-0085131P.



```
XX ADK48759;
AC
XX 20-MAY-2004 (first entry)
DT
XX Streptococcus pneumoniae protein, Seq ID No 5274.
DE
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
KW
XX Streptococcus pneumoniae.
OS
XX US6699703-B1.
PN
XX 02-MAR-2004.
PD
XX 26-MAY-2000; 2000US-00583110.
PF
XX 02-JUL-1997; 97US-0051553P.
PR
XX 12-MAY-1998; 98US-0085131P.
PR
XX 30-JUN-1998; 98US-00107433.
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
PI
XX WPI; 2004-212399/20.
DR
XX N-PSDB; ADK46098.
DR
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
XX Disclosure; SEQ ID NO 5274; 301pp; English.
XX
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
SQ
Query Match 99.6%; Score 796; DB 8; Length 2138;
Best Local Similarity 99.4%; Pred. No. 5e-70;
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEK 60
DB 1951 DTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEK 2010
QY 61 DAGVYINLSKDTFIKVPFKKIEEKEENKPTFDYKKNQPNVHNSQLNESHKREDLQR 120
DB 2011 DAGVYINLSKDTFIKVPFKKIEEKEENKPTFDYKKNQPNVHNSQLNESHKREDLQR 2070
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 2071 EDHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
AC AAW55096;
XX
XX 02-OCT-1998 (first entry)
DT
XX Streptococcus pneumoniae SP0043 protein.
DE
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
OS
XX WO9818930-A2.
PN
XX 07-MAY-1998.
PD
XX 30-OCT-1997; 97WO-US019422.
PF
XX 31-OCT-1996; 96US-0029960P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX WPI; 1998-272224/24.
DR
XX N-PSDB; AAV27357.
DR
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
SQ
Query Match 77.0%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKVPFKKIEEKEENKPTFDYSK 97
DB 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKVPFKKIEEKEENKPTFDYSK 60
QY 98 KKDNPQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 61 KKDNPQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
XX AC ABP54590;
XX
XX 04-SEP-2002 (first entry)
DT
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
DE
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
OS
XX US2002061545-A1.
PN
```



```

XX PD 23-MAY-2002.
XX PF
XX PA 22-JAN-2001; 2001US-00765272.
XX PA 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
Query Match 77.0%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 97
DB 1 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 98 KQNPQVNHSQLNSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
DB 61 KQNPQVNHSQLNSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX US6573082-B1.
XX PN 03-JUN-2003.
XX PD 28-MAR-2000; 2000US-00536784.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR

```

```

PR 30-OCT-1997; 97US-00961083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 58pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
Query Match 77.0%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 97
DB 1 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 98 KQNPQVNHSQLNSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 154
DB 61 KQNPQVNHSQLNSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 13
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX AC AAG81779;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX KW endocarditis.
XX OS Staphylococcus epidermidis.
XX XX WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.

```

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
XX  
XX Claim 18; Page 208; 2188pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AA881454 to AA883120, from *Staphylococcus epidermidis*. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce host cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
XX Sequence 746 AA;  
XX

Query Match	14.8%	Score 118;	DB 4;	Length 746;
Best Local Similarity	27.2%;	Pred. No. 0.014;		
Matches	47;	Conservative	23;	Mismatches 57; Indels 46; Gaps 9;
Qy	1	DTGVSSELKHRVTVTTIONGKEMSGSTTIVSEEDFILPVYK-----GELEKYQPDGM--- 51		
Db	586	DSVNAQSILKP-----ITINGKQIKQQSVKSTKVLPSPSKVLMLTGELTMP-DMTGWTKYE 640		
Qy	52	EISGFZ-----CKKDAGYVIN--LSKDTPIPKVPFKIEKKKEENKPTDVS-----KK 98		
Db	641	DVLAFEDLTKLVSTKGNGFVTNQSIKQIILK-----NKKDLVLSLAEDT 687		
Qy	99	KDNPNVNHSQLNESHRKEDLQREBHSQKSDSTKDVATVTLDKNNISSKSTNN 151		
Db	688	DDDOEKTDPSDDNKSKQKADSDHNTSSSTKN-----DKSNADSQNDSD 734		

RESULT 14	
ABP39023	
ID	ABP39023 standard; protein; 778 AA.
XX	
XX	ABP39023;
XX	
XX	24-JUL-2002 (first entry)
XX	
XX	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.
XX	
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX	antibacterial; gene therapy.
XX	
XX	Staphylococcus epidermidis.
OS	
XX	
PN	US6380370-B1.
XX	
XX	30-APR-2002.
PD	
XX	
XX	13-AUG-1998; 98US-00134001.
PF	
XX	
XX	14-AUG-1997; 97US-0055779P.
PR	
PR	08-NOV-1997; 97US-0064964P.
XX	
XX	(GENO-) GENOME THERAPEUTICS CORP.
PA	
XX	
PI	Doucette-Stamm LA, Bush D;

	WPI; 2002-381255/41.
N-PSDB	; ABN91568.
Novel isolated nucleic acid encoding a Staphylococcus epidermis polypeptide, useful for diagnosing and treating bacterial infections.	
Disclosure;	SEQ ID NO 3868; 267pp; English.
ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The S. epidermidis sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly S. epidermidis infections. The sequences can be used to screen for compounds able to interfere with the S. epidermidis life cycle or inhibit S. epidermidis infection. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site	
Sequence 778 AA;	
Query Match	14.8%; Score 118; DB 5; Length 778;
Best Local Similarity	27.2%; Pred.No. 0.015;
Matches	47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;
QY	1 DTGEVSLKPHRVTVTIQTONGEMSSIVSEEDFILPVYK-----GELEKGYYDGG--- 51   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	618 DSVNAQSLLKP----ITIGNGKQLIKQJVKSGTKVLPHSKVMMLTMDGELTMP-DMTGWTKE 672   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY	52 EISGFPE-----GKKDAGVVIN--LSKDTFIKPFVKKEEKEBENKPTFDYS-----KK 98   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	673 DVLAFLDLTKIYSTVGXGPVTNQGISKGQIIK-----NKOKIEVSLSAEDT 719   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY	99 KDNPNVNHSQNLRSHRKEDIQRHESKSQSDSTKVATATVLDKNNISKSSTNN 151   : :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	720 DDQEKTDEDSNDKSKOOKADEHDNTSSTTKN-----DKSNADSKNSDD 766   : :       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
RESULT 15	
ADS06368	
ID	AUS06368 standard; protein; 778 AA.
AC	ADS06368;
XX	
DT	04-NOV-2004 (first entry)
XX	
DE	Staphylococcus epidermis polypeptide seqid 5663.
KW	antibacterial; vaccine; antisease therapy; Staphylococcus epidermidis; recombinant expression vector; infection; computer readable medium; computer based system.  Staphylococcus epidermidis.
OS	Staphylococcus epidermidis.
PV	US2004147734-A1.
PD	29-JUL-2004.
PP	01-DEC-2003; 2003US-00724972.
PR	08-NOV-1997; 97US-0064964P.
PR	13-AUG-1998; 98US-00134001.
PR	29-NOV-1999; 99US-00450969.
PA	(DOUC/) DOUCETTE-STAMM L.
FA	(BUSH/) BUSH D.
PI	Doucette-Stamm L, Bush D;
DR	WPI; 2004-580138/56.
DR	N-PSDB; ADS02596.

PT New isolated polypeptide and encoding nucleic acid derived from  
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
 XX treating an S. epidermidis bacterial infection.  
 PS Claim 17; SEQ ID NO 5663; 741pp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for S. epidermidis infection; a recombinant or substantially  
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the Staphylococcus genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the Staphylococcus plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the Staphylococcus  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the Staphylococcus genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial  
 CC infection. This is the amino acid sequence of a S. epidermis protein of  
 XX the invention.

XX Sequence 778 AA;

Query Match 14.8%; Score 118; DB 8; Length 778;  
 Best Local Similarity 27.2%; Pred. No. 0.015; 57; Indels 46; Gaps 9;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW--- 51  
 Db 618 DSVNAQSLKP-----ITIGNGKIQKQSVKSTKVLPHSKVLMVMTDGLTMP-DMTGWTKE 672  
 QY 52 EISGFE-----GKKDAGYVIN--LSKDTPIKPVFKIEKKEENKPTFDVS-----KK 98  
 Db 673 DVLAFEDLTIKIVSTKNGFVTNQISKGQIIK-----NKDKIEVLSABDT 719  
 QY 99 KNPQVNHSQLNESHKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNN 151  
 Db 720 DDQBEKTDSDSNKSKDKDAEDHSNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 16  
 ABU42797  
 ID ABU42797 standard; protein; 775 AA.  
 AC ABU42797;  
 XX 19-JUN-2003 (first entry)  
 DE Protein encoded by Prokaryotic essential gene #28324.  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Staphylococcus epidermidis.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA46667.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 70721; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 775 AA;

Query Match 13.8%; Score 110.5; DB 6; Length 775;  
 Best Local Similarity 25.0%; Pred. No. 0.084;  
 Matches 42; Conservative 27; Mismatches 66; Indels 33; Gaps 8;  
 QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW--- 51  
 Db 615 DSVNAQSLKP-----ITIGNGKIQKQSVKSTKVLPHSKVLMVMTDGLTMP-DMTGWTKE 669  
 QY 52 EISGFE-----GKKDAGYVIN--LSKDTPIKPVFK-----KIEKKEENKPTFD 94  
 Db 670 DVLAFEDLTIKIVSTKNGFVTNQISKGQIIKNDKIEVLSABDTDDDEKTDSSD 729  
 QY 95 VSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDTATVLDKNN 142  
 Db 730 KSKKDKYDVEDNSNASSSSKNEKSNADSKNDSDBSTNETSGS--ERNN 775

RESULT 17  
 ADZ79639  
 ID ADZ79639 standard; protein; 188 AA.

XX ADZ79639;  
AC  
XX  
DT 14-JUL-2005 (first entry)  
DE  
XX P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.  
DE  
XX immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
KW immunotherapy; malaria; antimalarial; vaccine.  
XX  
XX Plasmodium falciparum.  
OS  
XX WO2005040206-A1.  
PN  
XX  
XX 06-MAY-2005.  
PD  
XX  
XX 22-OCT-2004; 2004WO-EP012910.  
PF  
XX  
XX 24-OCT-2003; 2003US-00691672.  
PR  
XX  
XX (INSP ) INST PASTEUR.  
PA  
XX  
XX Drulhe P;  
PI  
XX  
XX WPI; 2005-355821/36.  
DR  
XX  
XX Chimeric molecule useful for preparing vaccine composition against  
PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.  
XX  
XX  
XX Disclosure; SEQ ID NO 7; 79pp; English.  
XX  
XX The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
CC MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the  
CC Sequence Listing is not mentioned elsewhere in the specification.  
XX  
XX  
SQ Sequence 188 AA;  
Query Match 13.5%; Score 108; DB 9; Length 188;  
Best Local Similarity 23.2%; Pred. No. 0.021;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 17 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 69  
DB 15 VLKAKEASS-----YDYIL-----GWFGGGVPEHKKEENMLSHLYVSKD 55  
QY 70 KDTFKVPFKIEKKKEE-----ENKPTFDVSKKQDNPNQVHSQLNESHKRE 116  
DB 56 KENISKENDVDLDEKEEAETEELEKNEETSEISEDEEEEEEEKEEENDKKK 115  
QY 117 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 151  
DB 116 EQEKEQSNENNDDQKQMEA-----QNLISKQNNN 145

RESULT 18  
ADZ72253  
ID ADZ72253 standard; protein; 354 AA.  
XX  
XX ADZ72253;  
AC  
XX  
XX 14-JUL-2005 (first entry)  
DT  
XX Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.  
DE  
XX Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.  
XX  
XX Plasmodium falciparum.  
OS  
XX EP1526178-A1.  
PN  
XX  
XX 27-APR-2005.  
PD  
XX  
XX 24-OCT-2003; 2003EP-00292673.  
PF  
XX  
XX 24-OCT-2003; 2003EP-00292673.  
PR  
XX  
XX (INSP ) INST PASTEUR.  
PA  
XX  
XX Drulhe P;  
PI  
XX  
XX WPI; 2005-323987/34.  
DR  
XX  
XX N-PSDB; ADZ72252.  
XX  
XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
PT falciparum, which encode proteins useful for preparing vaccine  
PT compositions against malaria.  
XX  
XX Disclosure; SEQ ID NO 2; 137pp; English.  
XX  
XX The present invention relates to the protection against malaria. More  
CC particularly the invention pertains to a family of MSP-3 (merozoite  
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
CC falciparum, highly conserved in P. falciparum strains, simultaneously  
CC expressed in P. falciparum at the erythrocytic stages and encoding  
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
CC their N-terminal extremity and which are located at the merozoite  
CC surface. The characterization of this gene family enables the definition  
CC of immunogenic and vaccine compositions against P. falciparum. The  
CC present sequence is the P. falciparum MSP-3-1 protein.  
XX  
XX Sequence 354 AA;  
Query Match 13.5%; Score 108; DB 9; Length 354;  
Best Local Similarity 23.2%; Pred. No. 0.051;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 17 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 69  
DB 181 VLKAKEASS-----YDYIL-----GWFGGGVPEHKKEENMLSHLYVSKD 221  
QY 70 KDTFKVPFKIEKKKEE-----ENKPTFDVSKKQDNPNQVHSQLNESHKRE 116  
DB 222 KENISKENDVDLDEKEEAETEELEKNEETSEISEDEEEEEEEKEEENDKKK 281  
QY 117 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 151  
DB 282 EQEKEQSNENNDDQKQMEA-----QNLISKQNNN 311  
RESULT 19  
ADT56185  
ID ADT56185 standard; protein; 470 AA.  
XX  
XX ADT56185;  
AC

XX 13-JAN-2005 (first entry)  
XX Plant polypeptide, SEQ ID 6262.  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX Viridiplantae.  
XX US2004216190-A1.  
XX 28-OCT-2004.  
XX 18-DEC-2003; 2003US-00739930.  
XX 28-APR-2003; 2003US-00424599.  
XX 28-APR-2003; 2003US-00425115.  
XX (KOVA/) KOVALIC D K.  
XX Kovalic DK;  
XX WPI; 2004-757369/74.  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX Claim 2; SEQ ID NO 6262; 14pp; English.  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX Sequence 470 AA;

Query Match 13.5%; Score 107.5; DB 8; Length 470;  
Best Local Similarity 20.1%; Pred. NO. 0.084;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
QY 9 KPHRTVTIIONKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFE-----GKK 60  
DB 82 RENRVTDITVQNNNGESK-----YVQDLARRIRYDB-EATGQSAQRIDHPNQK 129  
QY 61 DAGVYVNLSDKDTFKPVFKIEKKKEENKPTFDVSKKKDN----- 101  
DB 130 NVGITERAFENSPIEETSHRVDDNKRINNQNFTAAKSSENAVSEVSGADHKRAEVNCK 189  
QY 102 POVNHSQLNE-----SHRKEDLQREHSQKSDSTKVDTATVLDKQNTSSKSTTNNPNK 154  
DB 190 PMENRDQVRQTESAEKSHRKENVTKSEKPRDQEGVKGKTEAKDKRNKKEKBEKTESINK 248  
RESULT 20  
AAG47777  
ID AAG47777 standard; protein; 484 AA.  
XX AC AAG47777;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 23-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127482P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 30-APR-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 18-MAY-1999; 99US-0134370P.  
XX 19-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145193P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.

PR 29-OCT-1999; 99US-0162142P.  
Query Match 13.5%; Score 107.5; DB 3; Length 484;  
Best Local Similarity 20.1%; Pred. No. 0.088;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
QY 9 KPHRVTVTTQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWEISGPE-----GKK 60  
DB 96 RENRVTDVTQNNNGESK-----YVQDLARRIRYDE-EATGSGSAQRIDHPNQK 143  
QY 61 DAGYVINLSKDTFIKPVFKKIEKEEENKPTFDVSKKDN-----101  
DB 144 NVGITEKAPENSPIETSHRVDDNKINNQNFTAAKSENNAVSRVSGADHRAEVMCK 203  
QY 102 POWNHSQLNE-----SHRKEDLQREHSQKSDSTKVDTATVLDKNISSTKTTNNPNK 154  
DB 204 PMENRDQVRQTSAAKSHRKNKVNKSEKPRDQGVKTEAKDKRNMKEKESKTESINK 262  
RESULT 21  
ADZ79635  
ID ADZ79635 standard; protein; 647 AA.  
XX AC ADZ79635;  
XX DT 14-JUL-2005 (first entry)  
XX DE P. falciparum GLURP-MSP3 fusion protein.  
XX KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;  
XX KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;  
XX KW vaccine.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2005040206-A1.  
XX PD 06-MAY-2005.  
XX PF 22-OCT-2004; 2004WO-EP012910.  
XX PR 24-OCT-2003; 2003US-00691672.  
XX PA (INSP ) INST PASTEUR.  
XX PI Druilhe P;  
XX DR WPI; 2005-355821/36.  
XX DR N-PSDB; ADZ79636.  
XX PT Chimeric molecule useful for preparing vaccine composition against  
PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.  
XX PS Disclosure; SEQ ID NO 3; 79pp; English.  
XX CC The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive

CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents plasmodium falciparum GLURP(27-  
CC 500)-MSP3(212-380) fusion protein.  
XX QY Sequence 647 AA;  
Query Match 13.3%; Score 106; DB 9; Length 647;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTTQNGKEM-----SSTIVSEEDFILPVYKGELEK-----44  
DB 428 ETVREHEETVSOESNPEKADNDGNVSNNNILNEFV-----ESEKSEHARSKAKEA 481  
QY 45 -GYOFD-QWETSGF--EGKKDAG-----YVINLSKDTFIKPVFKKIEKEEENKPTFDV 95  
DB 482 SSYDYLGNWFGGVPFHKKEENMLSHLYVSSKKENISKENDDVLDK-KEEEAEETEE 540  
QY 96 SKKDNPOVNHSQLN-----ESHKEDLQREHSQKSDSTKVDTATVLDKN 142  
DB 541 ELEKNEBETSEISEDEEBEKEEKEKEKEQKQKQKQNNNDQKQKMEA-----QN 595  
QY 143 ISSKSTTN 151  
DB 596 LISKNNNN 604  
RESULT 22  
AD019012  
ID AD019012 standard; protein; 651 AA.  
XX AC AD019012;  
XX DT 12-AUG-2004 (first entry)  
XX DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.  
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
XX KW antimalarial; immunostimulant.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2004043488-A1.  
XX PD 27-MAY-2004.  
XX PF 06-NOV-2003; 2003WO-DK000759.  
XX PR 12-NOV-2002; 2002DK-00001741.  
XX PR 11-SEP-2003; 2003DK-00001307.  
XX PA (STAT-) STATENS SERUM INST.  
XX PI Theisen M, Jepsen S;  
XX DR WPI; 2004-411650/38.  
XX PT New antigen based vaccine comprising a fusion protein derived from  
PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
PT preventing malaria.  
XX PS Disclosure; Fig 2C; 52pp; English.  
XX CC The present invention relates to a fusion protein comprising Plasmodium  
CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
CC useful as an antigen based vaccine against malaria. Also disclosed is the  
CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
CC polynucleotide sequence is also useful in preparing a vaccine. The



CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 hybrid.

XX SQ Sequence 651 AA;  
Query Match 13.3%; Score 106; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGLEK----- 44  
Db 432 ETVEHEETVSQESNPEKADNDGNVSNLNEFEV-----ESEKSEHARSKAKEA 485  
QY 45 -GYQPD-GWEISGP--EGKDGAG-----YVINLSKDTPIKPVFKIEKKEENKPTFDV 95  
Db 486 SSYDYLGMFEGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEE 544  
QY 96 SKKDNPPQVNHSQLN-----ESHKEDLQREHSQKSDSTKDVTTATVLDKNN 142  
Db 545 ELEKNEETESEISEDEEEEEEEKEEENEKKEQKESQNNNDQKQMEA-----QN 599  
QY 143 ISSKSTNN 151  
Db 600 LISKQNNN 608

RESULT 23  
ADO19010  
ID ADO19010 standard; protein; 651 AA.  
XX AC ADO19010;  
XX DT 12-AUG-2004 (first entry)  
XX DE P. falciparum GLURP-MSP3 fusion protein.  
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
XX KW antimalarial; immunostimulant.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2004043488-A1.  
XX PD 27-MAY-2004.  
XX PF 06-NOV-2003; 2003WO-DK000759.  
XX PR 12-NOV-2002; 2002DK-00001741.  
XX PR 11-SEP-2003; 2003DK-00001307.  
XX PA (STAT-) STATENS SERUM INST.  
XX PI Theisen M, Jepsen S;  
XX WPI; 2004-411650/38.  
XX DR N-PSDB; ADO19011.  
XX PT New antigen based vaccine comprising a fusion protein derived from  
XX PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
XX PT preventing malaria.

XX PS Claim 5; SEQ ID NO 1; 52pp; English.  
XX CC The present invention relates to a fusion protein comprising Plasmodium  
XX CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
XX CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
XX CC useful as an antigen based vaccine against malaria. Also disclosed is the  
XX CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
XX CC polynucleotide sequence is also useful in preparing a vaccine. The  
XX CC vaccine is useful in treating and preventing malaria and for inducing an

CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 fusion protein.

XX SQ Sequence 651 AA;

Query Match 13.3%; Score 106; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGLEK----- 44  
Db 432 ETVEHEETVSQESNPEKADNDGNVSNLNEFEV-----ESEKSEHARSKAKEA 485  
QY 45 -GYQPD-GWEISGP--EGKDGAG-----YVINLSKDTPIKPVFKIEKKEENKPTFDV 95  
Db 486 SSYDYLGMFEGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEE 544  
QY 96 SKKDNPPQVNHSQLN-----ESHKEDLQREHSQKSDSTKDVTTATVLDKNN 142  
Db 545 ELEKNEETESEISEDEEEEEEEKEEENEKKEQKESQNNNDQKQMEA-----QN 599  
QY 143 ISSKSTNN 151  
Db 600 LISKQNNN 608

RESULT 24  
AAB18278  
ID AAB18278 standard; protein; 665 AA.  
XX AC AAB18278;  
XX DT 07-NOV-2000 (first entry)  
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:135.  
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
XX KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX OS Plasmodium falciparum.  
XX PN WO200025728-A2.  
XX PD 11-MAY-2000.  
XX PF 05-NOV-1999; 99WO-US026796.  
XX PR 05-NOV-1998; 98US-0107131P.  
XX PA (HOFF/) HOFFMAN S.  
XX PA (CARU/) CARUCCI D.  
XX PA (GARD/) GARDNER M.  
XX PA (VENT/) VENTER J C.  
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX PT diagnosis of P.falciparum infection.  
XX PS Disclosure; Page 321-322; 577pp; English.

XX CC The present invention describes proteins and their fragments (I) encoded  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
XX CC vaccines against P. falciparum infection comprising (I) or (II); (1) and  
XX CC (II) are useful for the development of vaccines against P. falciparum  
XX CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
XX CC immunogens comprising the sequences of (I), are useful in the detection  
XX CC of infection with P. falciparum. Furthermore, (I) (especially when they  
XX CC are rifins or secreted or membrane proteins) can aid the identification

CC of drugs to treat or prevent P. falciparum infection, or they can be used  
 CC to identify drug resistance in P. falciparum. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAY70078 to AAA70287 and AAB18144 to AAB18152  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification  
 XX  
 SQ Sequence 665 AA;

Query Match 13.1%; Score 105; DB 3; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.24;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 QY 1 DTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI--SGPE 57  
 DB 128 EKNKINKSDLRHONELNLSQSGK-----NEQDI-----NKNEKGKQ----DISNSNAE 170  
 QY 58 GKQDAGYVNLKDTIFKPVFKKIEKKK-----EENKPTFD-----VSKKQNPQ 103  
 DB 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKKSDDHKVEENKKSDDHK 218  
 QY 104 VNHSQLNESHRKEDIQR-BEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 219 VEENKKSDDHKIEVKVKEHEDEEE-----DKKEKSEKNNKDENK 261

RESULT 25  
 ABO23606  
 ID ABO23606 standard; protein; 665 AA.  
 XX ABO23606;  
 DT 04-SEP-2003 (first entry)  
 XX Plasmodium falciparum outlier protein #3.  
 XX Candidate protein identification; pathogen; anti-infective;  
 KW outlier protein; virulence protein; antigen; drug target protein;  
 KW pathogenic organism; antimicrobial.  
 OS Plasmodium falciparum.  
 XX US2003039963-A1.  
 XX 27-FEB-2003.  
 XX 30-MAR-2001; 2001US-00820843.  
 XX 30-MAR-2001; 2001US-00820843.  
 XX (BRAH/) BRAHMACHARI S K.  
 PA (RAMA/) RAMACHANDRAN S.  
 PA (NAND/) NANDI T.  
 PA (BHIM/) BHIMARAO C.  
 XX Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;  
 XX WPI; 2003-492159/46.

Identifying candidate proteins useful as anti-infectives involves  
 PT matching outlier protein sequences with protein sequences in databases.  
 PT  
 PS Example 7; Page 91-93; 117pp; English.  
 XX The present invention relates to a method for identifying candidate  
 CC proteins in pathogens useful as anti-infectives. The invention discloses  
 CC a computational method which involves the calculation of several sequence

CC attributes and their subsequence analysis results in the identification  
 CC of outlier proteins in different pathogens. The method is useful for the  
 CC identification of outlier proteins (e.g. virulence proteins, antigens or  
 CC proteins used as drug targets) in pathogenic organisms. The method of the  
 CC invention provides reproducible results as it does not depend on the  
 CC variable biochemical characterisation of proteins. ABO23500-ABO23617  
 CC represent outlier proteins identified from different pathogenic organisms  
 XX  
 SQ Sequence 665 AA;

Query Match 13.1%; Score 105; DB 7; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.24;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 QY 1 DTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI--SGPE 57  
 DB 128 EKNKINKSDLRHONELNLSQSGK-----NEQDI-----NKNEKGKQ----DISNSNAE 170  
 QY 58 GKQDAGYVNLKDTIFKPVFKKIEKKK-----EENKPTFD-----VSKKQNPQ 103  
 DB 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKKSDDHKVEENKKSDDHK 218  
 QY 104 VNHSQLNESHRKEDIQR-BEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 219 VEENKKSDDHKIEVKVKEHEDEEE-----DKKEKSEKNNKDENK 261

RESULT 26  
 ADZ79634  
 ID ADZ79634 standard; protein; 169 AA.  
 XX ADZ79634;  
 DT 14-JUL-2005 (first entry)  
 XX P. falciparum merozoite surface protein 3, amino acid residues 212-380.  
 DE immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 KW immunotherapy; malaria; antimalarial; vaccine.  
 XX Plasmodium falciparum.  
 XX Key Location/Qualifiers  
 FH Region 1..169  
 FT /note= "Amino acid residues 212-380 of MSP3"  
 XX WO2005040206-A1.  
 XX 06-MAY-2005.  
 XX 22-OCT-2004; 2004WO-EP012910.  
 XX 24-OCT-2003; 2003US-00691672.  
 XX (INSP ) INST PASTEUR.  
 XX Druilhe P;  
 XX WPI; 2005-355821/36.

Chimeric molecule useful for preparing vaccine composition against  
 malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 immunized with molecule.  
 XX Claim 2; SEQ ID NO 2; 79pp; English.  
 XX The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also

CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3  
CC protein (amino acid residues 212-380).  
XX  
SQ Sequence 169 AA;

Query Match 13.0%; Score 103.5; DB 9; Length 169;  
Best Local Similarity 25.2%; Pred. No. 0.051;  
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
QY 21 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGF--EGKDGAG-----YVINLSKDTFF 73  
DB 1 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 41  
QY 74 IKPVFKIKIEKKEENKPTFVSKKDNQVNVHSLN-----ESHKEDLQR 120  
DB 42 SKENDVLDIE--KEEAEETEELEKKEEETTESISEDEEEEEEKKEEKKKEQEK 100  
QY 121 BEHSQKSDSTKDVATVLDKNNISSKSTTN 151  
DB 101 EQSNENNQKDMEA-----QNLISKNNNN 126

RESULT 27  
ID ABU25018  
AC ABU25018 standard; protein; 707 AA.  
XX ABU25018;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #10545.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Clostridium difficile.  
XX  
PN WO20027183-A2.  
XX  
PD 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342523P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA28888.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 52942; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: the sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 707 AA;

Query Match 12.9%; Score 103; DB 6; Length 707;  
Best Local Similarity 25.0%; Pred. No. 0.42;  
Matches 39; Conservative 25; Mismatches 56; Indels 36; Gaps 5;  
QY 1 DTGEVSELKPHRVTTTQNGKMSSTIVSEDFILPVYKGELEKGYQDFGWEISGFEGKK 60  
DB 504 DIGDVVEDKD-----TTDKYDVS---NKEDIIEPENKSKKKAKULFG----- 542  
QY 61 DAGVIVNLSKDTFIKPVFKIEKKEEN--KPTFDVSKKDNQVNVHSQLNESHKEDL 118  
DB 543 -----FIKONEEVEQEENLNDISPILLDKPVENNQVKSEETEQNLKE-I 589  
QY 119 QREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 154  
DB 590 KQEPSPQHIEBERSVKIEKPINNNLDEKVVSSNNEK 625

RESULT 28  
ABU61977  
ID ABU61977 standard; protein; 564 AA.  
XX  
AC ABU61977;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 12723.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX

PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEXE) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06080.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 564 AA;  
 SQ  
 Query Match 12.7%; Score 101.5; DB 4; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 0.43; Mismatches 29; Indels 25; Gaps 5;  
 Matches 34; Conservative 29;  
 QY 31 EDFILPVYKGELEKGYQFDG-----RISFEGKKGAGYVI-----NLSKDTFK 75  
 DB 78 ELDLPLSRSFSK--VFGWVDEHDEHGDHVDQEPSEALDDHDDHDDHDEDEE 135  
 QY 76 PVFKKLEKKEKREKPT-----FVSKKKQNPQVNSHLSHRKEDLQREHSHQKSDS 129  
 DB 136 PLTELELELEEEPTREDEPAADVEYEDDEENNA--GKNITAEAESEEBEDND 193  
 QY 130 TKDVTATVLDKNNISKST 148  
 DB 194 EGTVEATVEATTEATTEAT 212  
 RESULT 29  
 ABU25330  
 ID ABU25330 standard; protein; 1184 AA.  
 XX AC ABU25330;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #10857.  
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Clostridium difficile.  
 OS WO200277183-A2.  
 XX PN 03-OCT-2002.  
 XX PD 21-MAR-2002; 2002WO-US009107.  
 XX PF 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 26-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA29200.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 53254; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1184 AA;  
 SQ  
 Query Match 12.2%; Score 97.5; DB 6; Length 1184;  
 Best Local Similarity 26.3%; Pred. No. 3.1;  
 Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
 QY 4 EVSELKPHRVTVTIQ-NGKMSSTIV--SEEDP--ILPVYKGELEKGYQFDGWEISGPEG 58  
 DB 240 ELSRVNHRKVIKELNEKEKQKNVVEKKQDNKEVEVLQDVIEKSYDYN-SKGVIS 298  
 QY 59 KKDAGYVNLKSDTP-----IKPVFKKIEKKK-----EENKPTFDVSKKKD 100  
 DB 299 KKEE--QINLIKERTNFTNISRKNLEIKDKELNENKQVYKELESNK--LSGSEELS 354  
 QY 101 NPQVNSHLSHRKEDLQREHSHQKSDSTKDVATVLD-KNNISSKSTTNNPK 154  
 DB 355 TLQENIKVLEGSQKQKIKLSLNNELKESIIDILNKQKQEFNSKSLTLNANK 409  
 RESULT 30  
 ADP25441  
 ID ADP25441 standard; protein; 1791 AA.  
 XX AC ADP25441;  
 XX 09-SEP-2004 (first entry)  
 XX Plasmodium falciparum antigen amino acid sequence SEQ ID NO:18.



Qy	61	DAGVIVILSKDTFKVPVKIEE-----KREENKPTFDVSKKKONPOVNHSQLNESH	113
Db	1956	-----IEKDITTYKVKVENAVPAKQLKTKHN--TQSESQFKHTPVQKQOLVKYHN	2006
Qy	114	RKE--DLQREHSQKSDSTKDVTATVLDNNI	143
Db	2007	VKEQRSLEKGEHSTDMHVSELPETGETANKGL	2038

RESULT 32  
ADW88472  
ID ADW88472 standard; protein; 639 AA.  
XX  
XX AC AC ADW88472; /  
XX  
XX  
XX 21-APR-2005 (first entry)  
DT  
XX  
XX Staphylococcus aureus hybrid ORF0657n polypeptide.  
DE  
XX  
XX ORF0657n; vaccine; antibacterial; protein engineering;  
KW  
KW Staphylococcus aureus infection; mutein.  
XX  
XX  
XX Staphylococcus aureus.  
OS  
XX Synthetic.  
OS  
XX WO2005009378-A2.  
PN  
XX  
XX 03-FEB-2005.  
PD  
XX  
XX 22-JUL-2004; 2004WO-US023522.  
PF  
XX  
XX 24-JUL-2003; 2003US-0489840P.  
PR  
XX  
XX (MERI ) MERCK & CO INC.  
PA  
XX  
XX Anderson AS, Kuklin N, Jansen KU;  
PI  
XX WPI; 2005-123069/13.  
XX  
XX Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
PT useful for inducing protective immune response in humans against  
PT Staphylococcus aureus infection.  
XX  
XX Claim 7; SEQ ID NO 41; 84pp; English.  
XX  
XX The present sequence is that of a Staphylococcus aureus protein ORF0657n  
XX hybrid polypeptide; This is an example of claimed hybrid polypeptide  
XX immunogens of the invention ADW88439-ADW88474 that comprise a modified S.  
XX aureus ORF0657n sequence ADW88438 containing amino acid  
XX substitutions that increase sequence similarity to ORF0190 ADW88432. The  
XX hybrid polypeptides contain one or more epitopes for ORF0657n and  
XX ORF0190. They were designed by taking into account the similarity and  
XX differences between native ORF0657n and ORF0190 protein sequences. The  
XX invention also provides nucleic acids encoding these hybrid polypeptides,  
XX and a method for evaluating the ability of an immunogen to produce a  
XX protective immune response against Staphylococcus infection using an  
XX animal (mouse or rat) model. The hybrid polypeptides having therapeutic  
XX and diagnostic applications, such as being used to provide protective  
XX immunity against S. aureus infection, being used to generate antibodies  
XX to detect the presence of S. aureus, and being used to generate  
XX therapeutic antibodies that target S. aureus.  
XX  
XX Sequence 639 AA;  
SQ

QY	63	GYVINLSKDTFIKVPFKIIEKCKEENKPTDV-----SKKQDNQVNHSQLNESHKRD	11-
Db	445	YHVIRIVDKAEFTKANTDKSNKKGQDNSAKKEATPATPSKPTSPVKEKSRQDSQKDDN	504
QY	118	LQ-----REEHSQKSDSTKQVT-ATVLDKNNISSKSTTNNPNK	154
Db	505	KQLPSVEKENDASBSHGDKTPTAKPTKGEVSSSTT--PTK	544
RESULT 33			
AA	18272		
ID	AA518272	standard; protein; 2500 AA.	
XX	AC	AA518272;	
XX	AC		
DT	07-NOV-2000	(first entry)	
DE	Plasmodium falciparum	chromosome 2 related protein SEQ ID NO:129.	
XX			
KW	Plasmodium falciparum;	chromosome 2; human malaria parasite; vaccine;	
KW	antimalarial; malaria;	protozoacide; infection; insecticide.	
XX			
OS	Plasmodium falciparum.		
XX			
FN	WO200025728-A2.		
XX			
PD	11-MAY-2000.		
XX			
PF	05-NOV-1999;	99WO-US026796.	
XX			
PR	05-NOV-1998;	98US-0107131P.	
XX			
PA	(HOFF/) HOFFMAN S.		
PA	(CARU/) CARUCCI D.		
PA	(GARD/) GARDNER M.		
PA	(VENT/) VENTER J C.		
XX			
PI	Hoffman S, Carucci D, Gardner M, Venter JC;		
XX			
DR	WPI; 2000-365347/31.		

[illegible]

RESULT 34

RESOL 34  
ADS93954  
ID ADS93954 standard; protein; 635 AA.

AC ADS93954;

DT 02-DEC-2004 (first entry)

Fibrinogen-binding polypeptide, SEQ ID No 19.

KW fibrinogen-binding; adhesion factor; vaccine; bacterial infection;  
KW *Streptococcus agalactiae* infection; antibacterial; gene therapy;  
KW ribozyme; antisense; siRNA; anticalline; aptamer; Spiegelmer.

OS Streptococcus agalactiae.

PN WO2004035618-A2.

29-APR-2004.

PF 15-OCT-2003; 2003WO-EP011436.

PR 15-OCT-2002: 2002EP-00023141.

FR 20-MAR-2003; 2003EP-00006393.  
XX

PA (INTE-) INTERCELL AG.  
XX

PI Reinscheid DJ, Gutekunst H, Schubert A, Eikmanns BJ, Meinke A,   
yy

DR WPI; 2004-357201/33.

PT New nucleic acid molecules and encoded adhesion factors and/or fibrinogen  
PT -binding polypeptides for diagnosing, preventing or treating bacterial  
PT infections, preferably *Streptococcus agalactiae* infection.  
PT

PS Claim 13: SEO ID NO 19: 225pp: English.

The invention relates to a novel isolated nucleic acid molecule encoding a fibrinogen-binding polypeptide or its fragment, or an adhesion factor or its fragment. The invention further comprises: a vector comprising the above nucleic acid molecule; a cell, preferably a host cell, comprising the vector; a polypeptide, preferably a fibrinogen-binding polypeptide and/or an adhesion factor, comprising an amino acid sequence encoded by the above nucleic acid molecule; a process for producing the above polypeptide or its fragment; a process for producing a cell that expresses the polypeptide or its fragment; a pharmaceutical composition, especially a vaccine, comprising the polypeptide or its fragment, or the above nucleic acid molecule; an antibody, or its part, that binds to at least a selective part of the polypeptide or its fragment; methods for identifying an antagonist capable of reducing or inhibiting the activity of the polypeptide or its fragment, capable of binding to the polypeptide, or capable of reducing or inhibiting the interaction activity of the polypeptide or its fragment to its interaction partner; an antagonist identified by the above method; processes for in vitro diagnosis of a bacterial infection, preferably *Streptococcus agalactiae* infection, or a disease related to the expression of the above polypeptide or its fragment; and an affinity device comprising a support material and immobilized to the support material the above polypeptide or its fragment.



CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.

XX Sequence 635 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 635;  
 Best Local Similarity 24.5%; Pred. No. 2;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;

QY 19 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGFEKDKDAGYVI 66  
 DB 313 NTEPTLSYLENKEKFLVPNIPYKKNLILREEDKYSFEDDEBEFGNLLSYNKLKNEVLVPV 372

QY 67 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS-- 124  
 DB 373 NITTTILKP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQKVEQETKSPT 425

QY 125 QKSDSTEDVTATVLDKN--NISKSTT 149  
 DB 426 PQKTVKEQTEQKVSQNTQVEKKSET 452

# RESULT 36

ADV89902  
 ID ADV89902 standard; protein; 643 AA.

XX AC ADV89902;  
 XX DT 24-FEB-2005 (first entry)  
 XX DE Streptococcus agalactiae protein sequence, SEQ ID 2296.  
 XX KW Antibacterial; Vaccine; bacterial infection.  
 XX OS Streptococcus agalactiae.

XX FN FR2824074-A1.

XX PD 31-OCT-2002.

XX PF 26-APR-2001; 2001FR-00005642.

XX PR 26-APR-2001; 2001FR-00005642.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalloui L;  
 XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
 XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.

XX PS Claim 6; SEQ ID NO 2296; 2687pp; French.

XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;

CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transposons, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.

XX Sequence 643 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 2.1;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;

QY 19 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGFEKDKDAGYVI 66  
 DB 321 NTEPTLSYLENKEKFLVPNIPYKKNLILREEDKYSFEDDEBEFGNLLSYNKLKNEVLVPV 380

QY 67 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS-- 124  
 DB 381 NITTTILKP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQKVEQETKSPT 433

QY 125 QKSDSTEDVTATVLDKN--NISKSTT 149  
 DB 434 PQKTVKEQTEQKVSQNTQVEKKSET 460

# RESULT 37

ADV81155  
 ID ADV81155 standard; protein; 643 AA.

XX AC ADV81155;  
 XX DT 24-FEB-2005 (first entry)  
 XX DE Streptococcus agalactiae protein, SEQ ID 2296.  
 XX KW Antibacterial; vaccine; bacterial infection.  
 XX OS Streptococcus agalactiae.

XX FN WO200292818-A2.

XX PD 21-NOV-2002.

XX PF 26-APR-2002; 2002WO-IB003059.

XX PR 26-APR-2001; 2001FR-00005642.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Glaser P, Rusniok C, Chevallier F, Frangeul L, Lalloui L;  
 XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.

XX PS Claim 6; SEQ ID NO 2296; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and

CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
 CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.

XX  
 SQ Sequence 643 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 2.1;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 19 NKGKMSSTIVSEDFILP--VYKGELE-----EKGYQFD-----GWEISGFEGKKDAGYVIL 66  
 DB 321 NTEPLTSYLENKEKFLVPNIPYKNKLIILREEDKYSFEDDEBFNGNELLSYNKLKNEVLVPV 380  
 QY 67 NLSKDTFIKPVFKKTEEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS-- 124  
 DB 381 NITTTSTILKP-----FEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQVEQEKTKSPT 433  
 QY 125 QKSDSTKDVATVLDKN--NISSKSTT 149  
 DB 434 PQKRTVKEQTEQKVSNGTQVEKKSET 460

RESULT 38

ID ADW88460  
 XX ADW88460 standard; protein; 645 AA.  
 AC ADW88460;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 DE Staphylococcus aureus hybrid ORF0657n polypeptide.  
 XX  
 XX ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 XX  
 XX Staphylococcus aureus.  
 OS Synthetic.  
 XX  
 XX WO2005009378-A2.  
 FN  
 XX  
 XX 03-FEB-2005.  
 PD  
 XX  
 XX 22-JUL-2004; 2004WO-US023522.  
 PP  
 XX  
 XX 24-JUL-2003; 2003US-0489840P.  
 PR  
 XX  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX  
 XX Anderson AS, Kuklin N, Jansen KU;  
 PI  
 XX  
 XX WPI; 2005-123069/13.  
 DR  
 XX  
 XX Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.  
 XX  
 XX Claim 7; SEQ ID NO 29; 84pp; English.  
 PS  
 XX The present sequence is that of a Staphylococcus aureus protein ORF0657n  
 CC hybrid polypeptide. This is an example of claimed hybrid polypeptide

CC immunogens of the invention ADW88439-ADW88474 that comprise a modified *S.*  
 CC aureus ORF0657n sequence ADW88433-ADW88438 containing amino acid  
 CC substitutions that increase sequence similarity to ORF0190 ADW88432. The  
 CC hybrid polypeptides contain one or more epitopes for ORF0657n and  
 CC ORF0190. They were designed by taking into account the similarity and  
 CC differences between native ORF0657n and ORF0190 protein sequences. The  
 CC invention also provides nucleic acids encoding these hybrid polypeptides,  
 CC and a method for evaluating the ability of an immunogen to produce a  
 CC protective immune response against *Staphylococcus* infection using an  
 CC animal (mouse or rat) model. The hybrid polypeptides having therapeutic  
 CC and diagnostic applications, such as being used to provide protective  
 CC immunity against *S. aureus* infection, being used to generate antibodies  
 CC to detect the presence of *S. aureus*, and being used to generate  
 CC therapeutic antibodies that target *S. aureus*.

XX  
 SQ Sequence 645 AA;

Query Match 11.9%; Score 95; DB 9; Length 645;  
 Best Local Similarity 24.5%; Pred. No. 2.3;  
 Matches 49; Conservative 20; Mismatches 59; Indels 72; Gaps 10;  
 QY 16 TIQNGKMSSTIVSEDFILPVYKGELE-KGYQF-----DGWEISGFEGKKDAGYVILN 68  
 DB 355 SVENSESMOTFVEH-----PIKTGILNGKKYVMVMTNDDYKDFWVEGKR-----VRTI 405  
 QY 69 SK-----DTFIKPVFKKIB-----EKKEENKPTFDVSKKK 99  
 DB 406 SKDPKNTRTIIIPYVEGKALYDAIKVNVVKTIDYDQYHVRIVDKINTKANTDKSNKK 465  
 QY 100 D-----NPOVNHSQLNESHKEDLQ-----REEHSQKSDTKDVT- 134  
 DB 466 EQQNSAKKEATPATPSKPTSPVEKESQKDSQKDDNKQLPSVEKENDASSSESGKDKTP 525  
 QY 135 ATVLDKNNISKSSTNNPNK 154  
 DB 526 ATKPTGVESSSTT--PTK 543

RESULT 39

ID ADW88459  
 XX ADW88459 standard; protein; 645 AA.  
 AC ADW88459;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 DE Staphylococcus aureus hybrid ORF0657n polypeptide.  
 XX  
 XX ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 XX  
 XX Staphylococcus aureus.  
 OS Synthetic.  
 XX  
 XX WO2005009378-A2.  
 FN  
 XX  
 XX 03-FEB-2005.  
 PD  
 XX  
 XX 22-JUL-2004; 2004WO-US023522.  
 PP  
 XX  
 XX 24-JUL-2003; 2003US-0489840P.  
 PR  
 XX  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX  
 XX Anderson AS, Kuklin N, Jansen KU;  
 PI  
 XX  
 XX WPI; 2005-123069/13.  
 DR  
 XX  
 XX Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.  
 XX  
 XX Claim 7; SEQ ID NO 28; 84pp; English.  
 PS



**THIS PAGE BLANK (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 99.6107 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TVVKEFILNKDTGEVSELKP.....ATVLDKNNISKSTTNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Length	DB	ID	Description
1	848	100.0	2119	2	Q9AHT5_STRPN	Q9Aht5 streptococc
2	848	100.0	2140	2	Q97RY6_STRPN	Q97ry6 streptococc
3	845	99.6	2144	2	Q8DQP7_STRR6	Q8dqp7 streptococc
4	844	99.5	2144	2	Q9S4M8_STRPN	Q9s4m8 streptococc
5	133.5	15.7	300	2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	119	14.0	775	2	Q8CPK8_STRAP	Q8cpk8 staphylococ
7	113.5	13.4	296	2	Q5OLX8_ENTHI	Q5olx8 entamoeba h
8	112.5	13.3	361	2	Q95P15_PLAFA	Q95p15 plasmodium
9	112.5	13.3	379	2	Q9U6C4_PLAFA	Q9u6c4 plasmodium
10	111.5	13.1	346	2	Q9U0G0_PLARE	Q9u0g0 plasmodium
11	111.5	13.1	354	2	Q25995_PLARE	Q25995 plasmodium
12	111.5	13.1	354	2	Q8LJ55_PLAP7	Q8lj55 plasmodium
13	111.5	13.1	379	2	Q25706_PLAFA	Q25706 plasmodium
14	111.5	13.1	775	2	Q5HQ11_STRBQ	Q5hq11 staphylococ
15	111	13.1	1038	2	Q90784_CHICK	Q90784 gallus gall
16	110.5	13.0	379	2	Q25705_PLAFA	Q25705 plasmodium
17	110	13.0	500	2	Q8BGL7_PASTE	Q8bgl7 paramecium
18	110	13.0	829	2	Q8L5F3_PLAP7	Q8l5f3 plasmodium
19	109.5	12.9	384	2	Q5OVJ0_ENTHI	Q5ovj0 entamoeba h
20	109.5	12.9	609	2	Q8I2K8_PLAP7	Q8i2k8 plasmodium
21	109	12.9	380	2	Q26019_PLAFA	Q26019 plasmodium
22	109	12.9	3008	2	Q8I436_PLAP7	Q8i436 plasmodium
23	108.5	12.8	600	2	Q77355_PLAP7	Q77355 plasmodium
24	108.5	12.8	1069	2	Q5I2T7_ENTHI	Q5i2t7 entamoeba h
25	108	12.7	467	2	Q59PE2_CANAL	Q59pe2 candida alb
26	108	12.7	467	2	Q59PL2_CANAL	Q59pl2 candida alb
27	108	12.7	662	2	Q4YMU4_PLARE	Q4ymu4 plasmodium
28	107.5	12.7	470	2	Q9FJK9_ARATH	Q9fjk9 arabidopsis
29	107.5	12.7	1455	2	Q640L5_MOUSE	Q640l5 mus musculus
30	106.5	12.6	374	2	Q5V9M0_PLAKN	Q5v9m0 plasmodium
31	106.5	12.6	1015	2	Q5W5T1_TETPY	Q5w5t1 tetrahymena

32 106 12.5 616 2 Q6BRW2\_DEBHA Q6brw2 debaryomyce  
33 106 12.5 951 2 Q96229\_PLAP7 Q96229 plasmodium  
34 105.5 12.4 605 2 Q7RJC3\_PLAYO Q7rjc3 plasmodium  
35 105.5 12.4 674 2 Q7RL87\_PLAYO Q7rl87 plasmodium  
36 105 12.4 1550 2 Q54GS1\_DICDI Q54gs1 dictyosteli  
37 104.5 12.3 329 2 Q3NFV9\_PLAFA Q3nfv9 plasmodium  
38 104.5 12.3 736 2 Q4YVY2\_PLABE Q4yvvy2 plasmodium  
39 104 12.3 540 2 Q94CS9\_ARATH Q94cs9 arabidopsis  
40 104 12.3 540 2 Q9SA84\_ARATH Q9sa84 arabidopsis  
41 103.5 12.2 325 2 Q44016\_DICDI Q44016 dictyosteli  
42 103.5 12.2 393 2 Q7RKU2\_PLAYO Q7rku2 plasmodium  
43 103 12.1 2081 2 Q9LH98\_ARATH Q9lh98 arabidopsis  
44 103 12.1 2223 2 Q5TNJ2\_ANOGA Q5tnj2 anophelis g  
45 102.5 12.1 238 2 Q8I226\_PLAP7 Q8i226 plasmodium

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2S8T.  
DR MEROPS; S08.064; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR01899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR01680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Protease.

[illegible]

```

RESULT 2
Q97RY6_STRPN
ID Q97RY6_STRPN PRELIMINARY; PRT: 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641;
GS Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI TaxID=1313;

```

```

[1]
RN      NUCLEOTIDE SEQUENCE.
RP      STRAIN=ANCC BAA-334 / TIGR4;
RC      MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RX      Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA      Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA      Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA      Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA      Holtzapple E.K., Khouri H.M., Wolf A.M., Uterback T.R., Hansen C.L.,
RA      McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA      Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA      Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT      "Complete genome sequence of a virulent isolate of Streptococcus
      pneumoniae.";
RL      Science 293:498-506(2001).
DR      EMBL; AE007373; AAK74791.1; -; Genomic_DNA.
DR      PIR; F95074; F95074.
DR      HSSP; P00782; 2S8T.
DR      MEROPS; S08.064; -.
DR      TIGR; SP0641; -.
DR      GO; GO:0009986; C:cell surface; IEA.
DR      GO; GO:0005618; C:cell wall; IEA.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0008233; F:peptidase activity; IEA.
DR      GO; GO:0042802; F:protein self binding; IEA.
DR      GO; GO:0004289; F:subtilase activity; IEA.
DR      GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      InterPro; IPR010435; DUF1034.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      InterPro; IPR003137; PA.
DR      InterPro; IPR000209; Pept_s8_S53.
DR      InterPro; IPR010259; Prot_inh_S8A.
DR      InterPro; IPR001680; WD40.
DR      Pfam; PF06280; DUF1034; 1.
DR      Pfam; PF00746; Gram_pos_anchor; 1.
DR      Pfam; PF02225; FA; 1.
DR      Pfam; PF00082; Peptidase_S8; 1.
DR      Pfam; PF05922; Subtilisin_N; 1.
DR      PRINTS; PR00723; SUBTILISIN.

```

```

DR  TIGRFAMs; TTGR01167; LPXTG_anchor; 1.
DR  PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR  PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR  PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
DR  PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
KW  Cell wall; Complete proteome; Protease_.
SQ  SEQUENCE      2140 AA;  240426 MW;  F044D08E293B334 CRC64;

Query Match      100.0%; Score 848; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.4e-50;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  TTVKKEPTLNKDTGEVSELKPHRTVV;IQNGKEMSSITIVSEEDFILPVYKGELEKGYQFDG 60
      |||||
Db  1943  TTVKKEPTLNKDTGEVSELKPHRTVV;IQNGKEMSSITIVSEEDFILPVYKGELEKGYQFDG 2002
      |||||

Qy  61  WEISGFGKKDAGVYINSLKDTIKVFKKIEKKEENKPTPDVSKGKDNPOVNHSQLN 120
      |||||
Db  2003  WEISGFGKKDAGVYINSLKDTIKVFKKIEKKEENKPTPDVSKGKDNPOVNHSQLN 2062
      |||||

Qy  121  ESHRKEDLOREHSQKSDSTKQVTAFLDKNNISSTNNPNK 164
      |||||
Db  2063  ESHRKEDLOREHSQKSDSTKQVTAFLDKNNISSTNNPNK 2106
      |||||

```

RESULT 3	
Q8DQP7_STRR6	
ID	Q8DQP7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC	Q8DQP7;
DT	01-MAR-2003 (TREMBLrel. 23, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 26, Last annotation update)
DE	Cell wall-associated serine proteinase PrTA (EC 3.4.21.-).
GN	Name=prTA; OrderedLocusNames=spr0561;
OS	Streptococcus pneumoniae [strain ATCC BAA-255 / R6].
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=17101;
FN	[1]
RP	NUCLEOTIDE SEQUENCE.
RX	MEDLINE=21429245; PubMed=11544234;
RX	DOI=10.1128/JB.183.19.5709-5717.2001;
RA	Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burget
RA	Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Gerin
RA	Gilmour R., Glass J.S., Knoja H., Kraft A.R., Lagace R.E.,
RA	Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA	McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.
RA	Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA	Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA	Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud
RA	"Genome of the bacterium Streptococcus pneumoniae strain R6.";
RT	J. Bacteriol. 183:5709-5717 (2001).
RL	EMBL; AE008434; AAK9365.1; -; Genomic DNA.

```

DR  PIR: A97942; A97942.
DR  HSSP; P00782; 2SBT.
DR  MEROP; S08.064; -.
DR  GO; GO:0009986; C:cell surface; IEA.
DR  GO; GO:0005618; C:cell wall; IEA.
DR  GO; GO:0016020; C:membrane; IEA.
DR  GO; GO:0008233; F:peptidase activity; IEA.
DR  GO; GO:0042802; F:protein self binding; IEA.
DR  GO; GO:0004289; F:subtilase activity; IEA.
DR  GO; GO:0043086; F:negative regulation of enzyme activity; IEA.
DR  GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR  InterPro; IPR010435; DUF1034.
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  InterPro; IPR003137; PA.
DR  InterPro; IPR000209; Pept_S8_S53.
DR  InterPro; IPR010259; Prot_inh_S8A.
DR  InterPro; IPR001680; WD40.
DR  Pfam; PF06280; DUF1034; 1.
DR  Pfam; PF00746; Gram_pos_anchor; 1.

```

DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR Cell wall; Complete proteome.  
 KW SIGNAL 19 Potential.  
 FT CHAIN 20 2144 cell wall-associated serine proteinase  
 FT CHAIN 20 2144 PrtA.  
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 845; DB 2; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 3.9e-50;  
 Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFPIFKDGTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
 DB 1947 TTVKFPIFKDGTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 2006

QY 61 WEISGFEKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPPQVNHSQLN 120  
 DB 2007 WEISGFEKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPPQVNHSQLN 2066

QY 121 ESHRKEDLQREHDSQKSDSTKDVATVLDKNNISKSTTNNPNK 164  
 DB 2067 ESHRKEDLQREHDSQKSDSTKDVATVLDKNNISKSTTNNPNK 2110

RESULT 4  
 Q9S4M8 STRPN PRELIMINARY; PRT; 2144 AA.

AC Q9S4M8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2004 (TREMBlrel. 26, Last annotation update)  
 DE Cell wall-associated serine proteinase precursor PrtA.  
 GN Name=prtA;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3.B;  
 RX MEDLINE=21585565; PubMed=11728722;  
 RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,  
 RA Zysk G.;  
 RT "The cell wall-associated serine proteinase PrtA: a highly conserved  
 virulence factor of Streptococcus pneumoniae.";  
 RL FEMS Microbiol. Lett. 205:99-104(2001).  
 DR EMBL; AF127143; AAD48399.1; -; Genomic\_DNA.  
 DR HSSP; P00782; 2SBT.  
 DR MEROPS; S08\_064; -;  
 DR GO; GO:0009986; C:cell surface; IEA.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0042802; F:protein self binding; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0004308; F:negative regulation of enzyme activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR010435; DUF1034.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Pept\_S8\_SF3.  
 DR InterPro; IPR010259; Prot\_Inh\_SF8A.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF06280; DUF1034; 1.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.

DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR Cell wall; Signal.  
 KW SIGNAL 19 Potential.  
 FT CHAIN 20 2144 cell wall-associated serine proteinase  
 FT CHAIN 20 2144 PrtA.  
 SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 99.5%; Score 844; DB 2; Length 2144;  
 Best Local Similarity 98.8%; Pred. No. 4.6e-50;  
 Matches 162; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFPIFKDGTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
 DB 1947 TTVKFPIFKDGTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 2006

QY 61 WEISGFEKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPPQVNHSQLN 120  
 DB 2007 WEISGFEKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPPQVNHSQLN 2066

QY 121 ESHRKEDLQREHDSQKSDSTKDVATVLDKNNISKSTTNNPNK 164  
 DB 2067 ESHRKEDLQREHDSQKSDSTKDVATVLDKNNISKSTTNNPNK 2110

RESULT 5  
 Q4XUI6 PLACH PRELIMINARY; PRT; 300 AA.

AC Q4XUI6;  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Hypothetical protein (fragment).  
 GN ORFNames=PC000286.03.0;  
 OS Plasmodium chabaudi.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5825;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,  
 RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
 RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
 RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
 RA Bidwell S.B., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
 RA Jansse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.,  
 RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
 transcriptomic, and proteomic analyses.";  
 RL Science 307:82-86(2005).  
 CC -1- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC EMBL; CAJ01003049; CAH79425.1; -; Genomic\_DNA.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.  
 DR PRINTS; PR00320; GPROTEINERPT.  
 DR ProDom; PD0000018; WD40; 3.  
 DR SMART; SM00330; WD40; 5.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 4.  
 DR PROSITE; PS02094; WD\_REPEATS\_REGION; 1.  
 KW Hypothetical protein; Repeat; WD repeat.  
 FT NON\_TER 1  
 SQ SEQUENCE 300 AA; 34469 MW; 8ED84E512AFB1945 CRC64;

Query Match 15.7%; Score 133.5; DB 2; Length 300;  
 Best Local Similarity 26.6%; Pred. No. 0.18;  
 Matches 42; Conservative 32; Mismatches 51; Indels 33; Gaps 7;

QY 10 KDTGEVSELKPHRVTVVTIQNGKMSSTIVSE-----EDFILPVYKGELEKGYQPDG 61



```

Db      121 KSTRLLGHQKPVIRHQPSPNGKFIASSSPDKSIRIWSGIDGTYLAVYRHHVGPAYKI-AW 179
QY      62 EISGEGKKDAGYVNLNKDFTIK-----PVFKJIEEKE-----BENKPTFDVSKKGD 110
Db      180 SI-----DNNYIVSCSQDSTLKLWRINHLVPLLRKEENAEQTKDSQK-----NEQKE 227
QY      111 NFQVNHSQLNESHRRDLQREHSHOKSDSTKDVATVIL 148
Db      228 NFQ-NDQNDPNDSENEEKKKQKEKNDYTKNKTILL 264

RESULT 6
Q8CPK8 STAPR PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocusNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang H.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AE016746; AAO04453.1; -; Genomic_DNA.
DR HSSP; FL4677; IQME.
DR GO; GO:0008658; P:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001238; PBP_tpept_fold.
DR InterPro; IPR001460; Pencil_bind_tpept.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.0%; Score 119; DB 2; Length 775;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPHRVTVITQNGKEMSSIVSRSDFILPVYK-----GELEKGYQPDGW-- 61
Db 614 EDSVNAQSLKP-----ITGNHGQIKQSVKSTKVLPHSKVLMWTDGELTWP-DMTGWTK 668
QY 62 -EISGFE-----GKDDAGYVIN--LSKDTPIKPVFKKIEEKEENKPTFDVS-----K 107
Db 669 EDVLAFEDLTIKVSTKNGFVNTQSIKSGQIIK-----NKDKIEVLSAED 715
QY 108 KKDNPQVNHSQLNESHRRDLQREHSHOKSDSTKDVATVILDKNNISSKSTTN 161
Db 716 TDDQEKTDSDSNKSKKDAEDHSNTSSSTKN-----DKSNADSKNSDD 763

RESULT 7
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

```

```

DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amodeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW EMBL; AAPB01001439; EAL42595.1; -; Genomic_DNA.
SQ SEQUENCE 236 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;

Query Match 13.4%; Score 113.5; DB 2; Length 296;
Best Local Similarity 26.1%; Pred. No. 4.4;
Matches 41; Conservative 20; Mismatches 63; Indels 33; Gaps 5;

QY 4 KEFLNKDTGEVSELKPHRVTVITQNGKEMSSIVSRSDFILPVYKGELEKGYQPDGWEI 63
Db 135 KEQEKTKESGEGSEKKH--DIPTNEGKENTTKDKN-----KEEKDITVEEG-ES 184
QY 64 SGFGKDGAGVYVNLNKDTPIKPVFKKIEEKEENKPTFDVSKKDNPNVHSQLNESH 123
Db 185 SGKGGQTN-----ESKKTENPNQNESSNKKEEQ-----KKEEEK 221

RESULT 8
Q95PI5 PLAPA PRELIMINARY; PRT; 361 AA.
AC Q95PI5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=FVO;
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;
RA Hiseada H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,
RA Stowers A.W.;
RT "Merozoite surface protein 3 and protection against malaria in Aotus
RT nancymai monkeys.";
RL J. Infect. Dis. 185:657-664 (2002).
DR EMBL; AY044180; AAK94780.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT CHAIN
FT NON_TER 1 1
FT NON_TER 361 361 merozoite surface protein 3.

```

```
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 13.3%; Score 112.5; DB 2; Length 361;
Best Local Similarity 22.4%; Pred. No. 6.4;
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;

QY 11 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQ----- 57
Db 147 ETGE-----RNSRNFFYTKTKB-----YAGKVEKDYERAKNAYQKANQAV 187

QY 58 -----FD---GWEISGF--EGKKDAG-----YVNLKSDTFIKPVFKKIEKKKEEN 99
Db 188 LKAKRASSYDYLWGFEGGVPFHKKEENMLSHLYSSKDKENISKENDVDLDE-KEEA 246

QY 100 KPTFDVSKKQNPQVNHSLNHRKEDLQREHSOKSDSTKQVTVATVLDKN----- 151
Db 247 BETESEEKKEEETESISEDEBEKEEKEEENDKKQEKQESNENNDQKQDME 306

QY 152 --NISKSTNN 161
Db 307 AQNLISKNNNN 318

RESULT 9
Q9U6C4_PLAPA
ID Q9U6C4_PLAPA PRELIMINARY; PRT; 379 AA.
AC Q9U6C4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Polymorphic antigen.
GN Name=MSP-3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCCL/HN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match 13.3%; Score 112.5; DB 2; Length 379;
Best Local Similarity 23.6%; Pred. No. 6.8;
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 19 KPHRVTVTIQNGKMSSTIVSEEDF-----ILPVYKGELEKGYQFD-GWEISGF- 66
Db 171 KPSRLNLSRKTKEVAEQV--EKDYERAKNAYQKANQAVLKAKKASSYDYLWGFEGGV 228

QY 67 -EGKKDAG-----YVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNPVNHSLN 120
Db 229 PEHKKEENMLSHLYSSKDKENISKENDVDLDE-KEEAETETESEEKKEEETESIS 287

QY 121 -----ESHRKEDLQREHSOKSDSTKQVTVATVLDKNISKSTNN 161
Db 288 EDEBEKEEKEEKEEKKKEQEKQESNENNDQKQDME-----QNLISKNNNN 336

RESULT 10
Q9U0G0_PLARE
ID Q9U0G0_PLARE PRELIMINARY; PRT; 346 AA.
AC Q9U0G0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX NCBI_TaxID=5854;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON_TER 1 1
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAF010 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 346;
Best Local Similarity 23.9%; Pred. No. 7.2;
Matches 42; Conservative 32; Mismatches 67; Indels 35; Gaps 7;

QY 1 TTVKEPILN-KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD 59
Db 157 TKTKEVAQKAKNAYEKAKNAYQKANQAVLKAKKASS-----YNYIL----- 197

QY 60 GWEISGF--EGKKDAG-----YVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKQNP 112
Db 198 GWEFGGVPFHKKEENMLSHLYSSKDKENISKENDVDLDE-KEEAETETESEEKKEE 256

QY 113 QVNHSLNHRKEDLQREHSOKSDSTKQVTVATV-----LDKNISKSTNN 161
Db 257 BETESEEKKEEKEEKEEKEEENDKKQEKQESNENNDQKQDMEAQNLISKNNNN 312

RESULT 11
Q25995_PLAPA
ID Q25995_PLAPA PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RA McColl D.J., Silva A., Foley M., Kun J.P., Favaloro J.M.,
Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL EMBL; L28825; AAC09377.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 354;
Best Local Similarity 22.5%; Pred. No. 7.4;
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

QY 1 TTVKEPILN-KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD 59
Db 154 TKTKEVAQKAKNAYEKAKNAYQKANQAVLKAKKASS-----YDYIL----- 194
```

```

Qy 60 GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKKIEEKKEE-----EN 99
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 195 GWEEGGVPEHKKEENMLSHLYSSKKENISKENDVDLDEKEEAEETEEELKEKNEE 254
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 100 KPTFDVSKKNDPQNVHNSQLNESHKREDLQREHSQKSDSTKVDTATVLDKNNISKSTT 159
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 255 ETESEISEDEEEEEEEKEEENDKKQEKQESQNNNDQKKDKEA-----QNLISKQN 309
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 160 NN 161
    ||
Db 310 NN 311

RESULT 12
Q8IJ55_PLA77
ID Q8IJ55_PLA77 PRELIMINARY; PRT; 354 AA.
AC Q8IJ55;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN Q8FName:PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Portea M., Allen J., Selengut J.J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Vaidya A.B.,
RA McGadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF071133; Merozoite_SPAM; 1.
KW Merozoite.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 354;
Best Local Similarity 22.5%; Pred. No. 7.4;
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7

Qy 1 TTWKEFILN-KDTGEVSELKPHRVTTVTQNGKWSSTIVSEEDFILPVYKELEKGYQFD 59
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 154 TTKTKEAERAKNAVEKAKNAVQKANAQVLKAKEASS-----YDYIL----- 194
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 60 GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKKIEEKKEE-----EN 99
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 195 GWEEGGVPEHKKEENMLSHLYSSKKENISKENDVDLDEKEEAEETEEELKEKNEE 254
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 100 KPTFDVSKKNDPQNVHNSQLNESHKREDLQREHSQKSDSTKVDTATVLDKNNISKSTT 159
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 255 ETESEISEDEEEEEEEKEEENDKKQEKQESQNNNDQKKDKEA-----QNLISKQN 309
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 160 NN 161
    ||
Db 310 NN 311

RESULT 13
Q25706_PLAFA
ID Q25706_PLAFA PRELIMINARY; PRT; 379 AA.
AC Q25706;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

```

```
SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match
Best Local Similarity 13.1%; Score 111.5; DB 2; Length 775;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQPDGW-- 61
DB 614 EDSVNAQSLKP-----ITINGKQIKOOSVKSQGTGKVLPHSKVLMVMTDGLTMP-DMTGWTK 668
QY 62 -EISGFE-----GKDDAGYVIN--LSKDTTIKPVFK-----KIBEEKEENKPTF 103
DB 669 EDVLAPEDELTKIKVSTKNGFVTNQISKGQIKKLEVSLSAEDTDDQEKTDDESS 728
QY 104 DVSKKQNPQVNHSHRKEDELRQREHSQKSDTKDVTATVLDKNN 152
DB 729 DKSKKDKVDDEDSNASSSSKNEKSNADSKNDSSTNETSGS--ERNN 775

RESULT 15
ID Q90784_CHICK PRELIMINARY; PRT; 1038 AA.
AC Q90784;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Clausturin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Clausturin, an antiadhesive neural keratan sulfate proteoglycan, is
structurally related to MAP18."
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -; mRNA.
DR PIR; JC5497; JC5497.
DR Ensemble; ENSGALG0000014999; Gallus gallus.
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match
Best Local Similarity 13.1%; Score 111; DB 2; Length 1038;
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

QY 3 VKEFILKDTGEVSELKPHRVTVTIQN----- 29
DB 528 VQQAQKLQRTDSKESLKPAAKTTTKQDCQKRNKKKHSLSQSLVQOLEKPKLESKEKTP 587
QY 30 -----GKEMSTIVSEEDFILPVYKGELEKGYQPDGWEISFEGKDKAGYVINLSKDTF 83
DB 588 VKKKAQVPEPTIKYVAEKDV-----TTKEQLGKSETSKQASEKQDVKPKVTKRS 639
QY 84 IKPVFK-KLEBKEENKPTFDVSKKQNPQVNHSHRKEDELRQREHSQKSDTKDVTATVLDKNN 137
DB 640 VKKKAQVPEPTIKYVAEKDV-----TTKEQLGKSETSKQASEKQDVKPKVTKRS 137
QY 138 DSTKDV 143
DB 697 EAKKEV 702

RESULT 16
ID Q25705_PLAFA PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
```

```
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08851; AAC47831.1; -; Unassigned_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match
Best Local Similarity 13.0%; Score 110.5; DB 2; Length 379;
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;

QY 11 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQ----- 57
DB 167 ETGE-----ENSRNFFVTKTKB-----YAGKVKDYERAKNAYQKQNAV 207
QY 58 -----FD---GWEISGP--EGKQDAG-----YVINLSKDTFIKPVFKIIEKKBEEN 99
DB 208 LKAKEASSDYILGWFGGVPVPEHKKENMLSHLYVSSKDKENISKENDVDLDE-KBEEA 266
QY 100 KPTFDVSKKQNPQVNHSHRKEDELRQREHSQKSDTKDVTATVLDKNN 146
DB 267 BETEELHLEKKEETSEISEDEEBEKEEKEEKKKQKQKQENNDQKQKMEA- 325
QY 147 VLDKNNISSKSTTNN 161
DB 326 ----QNLISKQNNN 336

RESULT 17
ID Q6BGL7_PARTE PRELIMINARY; PRT; 500 AA.
AC Q6BGL7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PTMB_06c;
OS Paramacium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramacium.
OX NCBI_TaxID=5888;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Stock d4-2;
RX PubMed=15296759; DOI=10.1016/j.cub.2004.07.029;
RA Zagulski M., Nowak J.K., Le Mouel J.K., Nowacki M., Migdalski A.,
Gromadka R., Noel B., Blanc I., Dessen P., Wincker P., Keller A.M.,
Cohen J., Meyer B., Sperling L.;
RT "High Coding Density on the Largest Paramacium tetraurelia Somatic
Chromosome.";
RL Curr. Biol. 14:1397-1404(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Stock d4-2;
RA Nowak J.K., Migdalski A., Gromadka R., Zagulski M.;
RT "Paramacium megabase sequencing project.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR548612; CAH03203.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 500 AA; 56364 MW; DB4D7F90C86B79F0 CRC64;

Query Match
Best Local Similarity 13.0%; Score 110; DB 2; Length 500;
Matches 43; Conservative 34; Mismatches 75; Indels 40; Gaps 6;
```

```
QY 4 KEFILNKOTGEVSELKPHRVTTVITQNGKEMSS-----TIVSEEDFILPVYKG 50
Db 11 KQDLUKKAKEIRE-EFFQTVPIQNAESEKKLKMKVPNEHMTATSDDKILIDATLG 69
QY 51 ELEKGQFQGWISGFEKGKADGYVIN-----LSKDTFIKPVFK-----KIEBK 95
Db 70 NEEAQREVQIYQITPANQVRSYALTQEGVPVQRPDDFEVEMFKSPQMDKINLKIEKR 129
QY 96 EENKPTFPVSK---KQNPQNHLSQNSHRK-----BDLQREHSQKSDSTKDV 144
Db 130 DEKNKEEEMKHKSKLKNLQMKIKMREBEHKEKQNKVAIQWKKIEKSGDKARDLD 189
QY 145 ATVLDKNNISSK 156
Db 190 EIIKQNNQISKK 201

RESULT 18
ID Q815F3_PLAF7 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF1275c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Carlson J.M., Pain A., Fung E., Tamaki T., Wang F., Davis R.W.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS014848; AAN36341.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98815 MW; EF2675E301B2CE93 CRC64;

Query Match 13.0%; Score 110; DB 2; Length 829;
Best Local Similarity 24.0%; Pred. No. 23;
Matches 43; Conservative 39; Mismatches 55; Indels 42; Gaps 9;

QY 2 TVKEPTLNKDTGVSELKPHRVTTVITQNGKEMSSITVSEEDFILPVYKGELEKGYQFQGW 61
Db 491 TSNDSILNTKTKI--IQP--LEYLKNIGD--KTLMTERDIVLDFPHYMIKKYHLN-- 542
QY 62 EISGFEKGKADGYVINLSKDTFTIKPVFKIEKKEENKPT-----FDVSK 107
Db 543 -----KKEITLNFSLN-----FRIEKNRDKKKGTHNNKNDAEYMLKYIKK 587
QY 108 KQCN-PQVNHLSQNSH-RKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164
Db 588 KQCNFENNNTLNDNSIKENNKVLEH---DNSLKQEQIINDRKNVIEHTKIYDNQKK 643

RESULT 19
Q50VJ0_ENTHI
```

```
Q50VJ0_ENTHI PRELIMINARY; PRT; 384 AA.
Q50VJ0;
AC Q50VJ0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE HMG box protein.
GN ORFNames=188.t00012;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichevitz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAPB01000585; EAL45607.1; -; Genomic DNA.
SQ SEQUENCE 384 AA; 45464 MW; 207789F65D72B019 CRC64;

Query Match 12.9%; Score 109.5; DB 2; Length 384;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 39; Conservative 28; Mismatches 64; Indels 13; Gaps 4;

QY 17 ELKEHRRVTVITQNGKEMSSITVSEEDFILPVYKGELEKGYQFD--GWEISGFEKGKADGY 74
Db 226 EERTKKYVEIKEDDEKTKYVEIKED-----EKKEKKSKKEDKKKEEMKQNEKKSDDK 280
QY 75 VINLSKDTFTIKPVFKIEKKEENKPTFDVSKKQNPQVNHLSQNSHNRKEDLQREHS 134
Db 281 KEDTKKKKKVKKSEKKEIKKEDEKHKH---EKKEKTEKKPKKPSSEKSEKKEKKS 336
QY 135 QKSDSTKD--VTATVLDKNNISSK 156
Db 337 KKEDKKKDEKSKKVEDKSKKQK 360

RESULT 20
Q812K8_PLAF7 PRELIMINARY; PRT; 609 AA.
ID Q812K8_PLAF7 PRELIMINARY;
AC Q812K8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 26, Last annotation update)
DE Large cyclophilin-like protein.
GN Name=PF11490c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
```

RA	Knighte A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA	Line A., Maddison M., Mclean J., Money P., Moule S., Murphy E.,
RA	Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA	Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA	Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA	Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA	Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT	"Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL	Nature 419:527-531(2002).
DR	EMBL; AL329358; CAD51984.1; -; Genomic DNA.
DR	GO; GO:0006457; P-protein folding; IEA.
DR	InterPro; IPR002130; CSA_PPIase.
DR	PROSITE; PS00072; CSA_PPIASE_2; 1.
SQ	SEQUENCE 609 AA; 72551 MW; 8CDF86E85FF9A021 CRC64;
Query Match 12.9%; Score 109.5; DB 2; Length 609;	
Best Local Similarity 23.9%; Pred. No. 18;	
Matches 42; Conservative 36; Mismatches 79; Indels 19; Gaps 5;	
QY	5 EFILNKDTGVSELYKPHRVTVTTIQNGKEMSSTI-----VSEEDPFL----PVYKGSLERKYQ 57
DB	114 KIILNRKTGNVSYSEIYKL---FHNKNMPPDITHDKISKEDFVLQDLPLNNKTSYYVASH 170
QY	58 FDGWEISGPCKDAGVYNLSKDTFKVPVKLTBEKKKEENKTFPDVSKKKNP----- 112
DB	171 LNDRDQEDNEKKCNKIIFAPSIQDNKAIMSLKENFEAKRKRKEQESEEKCKKVGYIL 230
QY	113 QVNHSQLNSHRKEDLQREHESQKSDTKVTATVLD----KNISSKSTTNNPK 164
DB	231 QDYNSDSNEHSHSEKKNMKLNKKDEYKNKHSTNYSDSENHKIKCTYSDNK 286
RESULT 21	
Q26019_PLAPA	
ID	Q26019_PLAPA PRELIMINARY; PRT; 380 AA.
AC	Q26019
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
OS	Polymorphic antigen precursor.
DE	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;	[1]
RN	NUCLEOTIDE SEQUENCE.
RC	STRAIN=PC27;
EX	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA	McColl D.J., Silva A., Foley M., Kun J.F., Favalaro J.M.,
RA	Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT	"Molecular variation in a novel polymorphic antigen associated with
RT	Plasmodium falciparum merozoites";
RL	Mol. Biochem. Parasitol. 68:53-67(1994).
RP	[2]
NUCLEOTIDE SEQUENCE.	
RC	STRAIN=PC27;
RX	MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA	McColl D.J., Anders R.F.;
RT	"Conservation of structural motifs and antigenic diversity in the
RT	Plasmodium falciparum merozoite surface protein-3 (MSP-3)."
RL	Mol. Biochem. Parasitol. 90:21-31(1997).
DR	EMBL; L07944; AAC09378.1; -; Genomic_DNA.
DR	PDB; 1PSM; NMR; @=90-127.
DR	InterPro; IPR010784; Merozoite_SPAM.
DR	Pfam; PF07133; Merozoite_SPAM; 1.
KW	Signal.
FT	SIGNAL 1 25 Potential.
FT	CHAIN 26 380 polymorphic antigen.
SQ	SEQUENCE 380 AA; 43290 MW; 0586CA1393094CA2 CRC64;

Query Match 12.9%; Score 109; DB 2; Length 380;  
Best Local Similarity 23.5%; Pred. No. 12;  
Matches 42; Conservative 34; Mismatches 67; Indels 36; Gaps 8;

Qy	12	TGVESELKPHRVTVTTQNGKEMSGTIVSBDFFILPVY-----KGELEKGYQFD-GWE	62
Db	166	TSETPE-KPSRIINLFKSKTKYEAKAKNAVEKAKNAYQKANAQAVLKAKEASSYDYLIGE	224
Qy	63	ISGF--EGKQDAG-----YVINLSKDTFFIKPVFKKIEKKKEENKPTFDVSKKKDNPQVN	115
Db	225	FGGVPFHKKCKENMLSHLYVSSKDKENISKENDVDVLDE-KKEABETEEBELEKNEBET	283
Qy	116	HSOLN-----BSHRKEDLQREHSOKSDSTKDVTVTLVDKKNISKSSTNN	161
Db	284	ESISEDEBEEREEKEENKKEKQEKQESNIENNQKKDMEA-----QWLISKQNNN	337
RESULT 22			
Q81436_PLAF7			
ID	Q81436	PLAF7 PRELIMINARY; PRT; 3008 AA.	
AC	Q81436		
DT	01-MAR-2003	(TEMBLrel. 23, Created)	
DT	01-MAR-2003	(TEMBLrel. 23, Last sequence update)	
DT	01-MAR-2003	(TEMBLrel. 23, Last annotation update)	
DE	Hypotheical protein PF0325w.		
GN	Name=PF0325w;		
OS	Plasmodium falciparum (isolate 3D7).		
OC	Rukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
NCBI_TaxID=	36329;		
RN	[1]		
RC	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=3D7;		
EX	MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;		
RA	Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,		
RA	Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,		
RA	Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,		
RA	Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,		
RA	Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,		
RA	Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,		
RA	Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,		
RA	Humphray S., Jagels K., James K.D., Johnson D., Kethornou A.,		
RA	Knights A., Konfortov B., Kyes J.S., Larke N., Lawson D., Lennard N.,		
RA	Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,		
RA	Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,		
RA	Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,		
RA	Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,		
RA	Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,		
RA	Sulston J.R., Craig A., Newbold C., Barrell B.G.;		
RT	"Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";		
RV	Nature 419:527-531(2002).		
RN	[2]		
RC	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=3D7;		
RA	Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,		
RA	Hall N., Bowman S., Churcher C., Quail M., Barrell B.;		
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL929351; CAD51431.1; -; Genomic_DNA.		
KW	Hypotheical protein.		
Qy	SEQUENCE 3008 AA; 356025 MW; 60BCBBEE15C599B4 CRC64;		
Db	Query Match 12.9%; Score 109; DB 2; Length 3008;		
Db	Best Local Similarity 32.4%; Pred.No. 1.1e+02;		
Db	Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;		
Qy	74	YVINLSK----DTFFIKPVFKKIEKKKEENKPTFDVSKKKDNPQVHQS---LNESHKKE	126
Db	2310	YDIELSKIEKFGASIGPVFTD-EENKKEEN--EVNKKENKCKEENKNEVNIKE	2366
Qy	127	LQREH-----SKSDSTKDVTVTLVDKKNISK-----STNNPNK	164
Db	2367	ENKKEENKCKEENKCKEENKNEVNIKEENKCKEENKCKEENKCKEENKCKEENKCKEENK	2414
RESULT 23			
O77355_PLAF7			
ID	O77355	PLAF7 PRELIMINARY; PRT; 600 AA.	
AC	O77355		

RESULT 23  
O77355 PL  
ID O773  
AC O773



```

DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein MAJ3P4.20
GN Name=MAJ3P4.20; Synonym=PF00465c;
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jaesal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RA "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Corton C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Cronin A.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lacke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAA15610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F-lyase activity; IEA.
DR GO; GO:0006397; P:RNA processing; IEA.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;
Query Match 12.8%; Score 108.5; DB 2; Length 600;
Best Local Similarity 29.3%; Pred. No. 21;
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;
QY 37 IVSEDFILPVY-----KGELEKGYQDGEWISGEGKK----DAGYVNLKDTPIKPV 87
DB 60 ILGFDDIILYECISQLKQSKKK---DGEEDKYNLAKLNTLGTIGNKKSDFIBEL 116
QY 88 FKKI--EEKKEE-----ENKPTFDVSK-KKDNPOVHNSQLNE-----SHRK 125
DB 117 LELLNEEKKKEHIADTLNENK-TNDIKKVKVENENINENVNENKDNISNKKDKEHVSHQN 175
QY 126 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISKSTTN 160
DB 176 EHNINNVNLKKEKEYTDIQDKRKHKRLSKQSDSYKKPFPNKRKTSIER-SLSNRYDE 234
QY 161 NPNK 164
DB 235 KTNK 238

```

```

RESULT 24
Q512T7_ENTHI PRELIMINARY; PRT; 1069 AA.
AC Q512T7_
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J., Nozaki T.,
RA Amodeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sacheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RA "The genome of the protist parasite Entamoeba histolytica.";
RT Nature 433:865-868(2005).
RL CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAPB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;
Query Match 12.8%; Score 108.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 39;
Matches 42; Conservative 34; Mismatches 60; Indels 29; Gaps 7;
QY 14 EVSELKPRVTVTTQNG-KEMSSIVSEEDFILPVKGELEKGYQDGEWISGEGKKDA 72
DB 138 EVSKDNVNSSSTLTNGEKKLSTSLCNEQD-----ELQKSSSSSTD--NKNDRKDE 187
QY 73 GYVNLKSDTPIKPVFKIEEKEENK-----PTFDVSKKKDNPQ-----VNHSQLN 120
DB 188 IHFVDVLPKNEEISMEIESKTEEEKSNLIQISLNLSECKDNESVEIAKVLKSNSS 247
QY 121 ESHRKEDLQREHS-QKSDSTKD-----VTATVLDKNNISKSTT 159
DB 248 NNSGEEDKQDEVSCEKFDQSBEKKEEMIKAEVSNKQEVKDKSTT 292
RESULT 25
Q59PE2_CANAL PRELIMINARY; PRT; 467 AA.
AC Q59PE2_
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=Cao19.6351;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA Davis R.W., Scherer S.;
RA "The diploid genome sequence of Candida albicans.";
RT

```



RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).

RN [2] NUCLEOTIDE SEQUENCE.  
 RC STRAIN=SC5314;  
 RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,  
 RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,  
 RA Jones T., Scherer S., Agabian N.;  
 RA "Annotation of the Genome of Candida albicans."  
 RT Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AACQ01000192; EAK92345.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 467 AA; 52829 MW; 2F4D37A2127A7253 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 467;  
 Best Local Similarity 20.8%; Pred. No. 17;  
 Matches 50; Conservative 36; Mismatches 58; Indels 96; Gaps 9;

QY 10 KDTGEVSEL-KPHRVTVT-----IQNGKMSSTIVSEEDP----- 43  
 DB 189 KSTPKTSPLRKPKPTVTPVRKMAKRPSPATNTPETPKKSSSPPIISESDFLEMD 248  
 QY 44 ----ILPVYKGELEKGYQDGEWISGFEKGDAGVYVNLKDTFKIPVKPKIEKKEEN 99  
 DB 249 KSTERVPIIE-----FNFNDYD--NDEDKKEE--VVKSKNE--NQNITKGMESKPKPK 297  
 QY 100 KPTFDVSKKKNPQVN-----HSQLNESHKHEKDLQREE 132  
 DB 298 KPESEATKTKVQPKQKQKPLSEETVLTDDDDFKDLEQLEELLEEEBQPKQ 357  
 QY 133 HSQKSDSKDV-----TATVLDKNMISSTNNPNK 164  
 DB 358 QIKETKSNQSIQKPKSPPTIEVDPIAFNDSDFEDFHTGKIDEGNNSSSNNNK 417

# RESULT 26

Q59PL2 CANAL  
 ID Q59PL2 CANAL PRELIMINARY; PRT; 467 AA.  
 AC Q59PL2  
 DT 10-MAY-2005 (TREMELrel. 30, Created)  
 DT 10-MAY-2005 (TREMELrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TREMELrel. 30, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=Ca019.13708;  
 OS Candida albicans SC5314.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=237561;  
 RN [1] NUCLEOTIDE SEQUENCE.  
 RC STRAIN=SC5314;  
 RX PubMed=15123810; DOI=10.1073/pnas.0401648101;  
 RA Jones T., Federispiet N.A., Chibana H., Dungan J., Kalman S.,  
 RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,  
 RA Davis R.W., Scherer S.;  
 RT "The diploid genome sequence of Candida albicans."  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).  
 RN [2] NUCLEOTIDE SEQUENCE.

RC STRAIN=SC5314;  
 RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,  
 RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,  
 RA Jones T., Scherer S., Agabian N.;  
 RA "Annotation of the Genome of Candida albicans."  
 RT Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AACQ01000189; EAK92416.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 467 AA; 52769 MW; CFBES61B6BD8B588 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 467;  
 Best Local Similarity 20.8%; Pred. No. 17;  
 Matches 50; Conservative 36; Mismatches 58; Indels 96; Gaps 9;

QY 10 KDTGEVSEL-KPHRVTVT-----IQNGKMSSTIVSEEDP----- 43  
 DB 189 KSTPKTSPLRKPKPTVTPVRKMAKRPSPATNTPETPKKSSSPPIISESDFLEMD 248  
 QY 44 ----ILPVYKGELEKGYQDGEWISGFEKGDAGVYVNLKDTFKIPVKPKIEKKEEN 99  
 DB 249 KSTERVPIIE-----FNFNDYD--NDEDKKEE--VVKSKNE--NQNITKGMESKPKPK 297  
 QY 100 KPTFDVSKKKNPQVN-----HSQLNESHKHEKDLQREE 132  
 DB 298 KPESEATKTKVQPKQKQKPLSEETVLTDDDDFKDLEQLEELLEEEBQPKQ 357  
 QY 133 HSQKSDSKDV-----TATVLDKNMISSTNNPNK 164  
 DB 358 QIKETKSNQSIQKPKSPPTIEVDPIAFNDSDFEDFHTGKIDEGNNSSSNNNK 417

# RESULT 27

Q4YMU4 PLABE  
 ID Q4YMU4 PLABE PRELIMINARY; PRT; 662 AA.  
 AC Q4YMU4;  
 DT 13-SEP-2005 (TREMELrel. 31, Created)  
 DT 13-SEP-2005 (TREMELrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMELrel. 31, Last annotation update)  
 DE RNA binding protein, putative.  
 GN ORFNames=PB001104.03.0;  
 OS Plasmodium berghei.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=5821;  
 RN [1] NUCLEOTIDE SEQUENCE.  
 RP NUCLEOTIDE SEQUENCE.  
 RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,  
 RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
 RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
 RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
 RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
 RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
 RA "A comprehensive survey of the Plasmodium life cycle by genomic,  
 RT transcriptomic, and proteomic analyses."  
 RL Science 307:82-86 (2005).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC EMBL; CAJ101003467; CAJ100666.1; -; Genomic\_DNA.  
 DR InterPro; IPR012972; NLE.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF08154; NLE; 1.  
 DR Pfam; PF00400; WD40; 8.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR ProDom; PD000018; WD40; 4.  
 DR SMART; SM00320; WD40; 8.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 662 AA; 75536 MW; E7C8543AF5E59124 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 662;  
 Best Local Similarity 25.6%; Pred. No. 25;  
 Matches 43; Conservative 28; Mismatches 55; Indels 42; Gaps 8;

QY 10 KDTGEVSELKPHRVTVTITQNGKMSSTIVSE-----EDFILPVYKGELEKGYQD 61  
 DB 472 KTKLLGHQKPVHTQFSPNGKFIASSFDKSIKWSIDGAYLAVFRGHVGPAYKI-AW 530  
 QY 62 EISGFBGKQDAGYVNLKDTFK-----PVFKKIE-----KKEENK-PTF 103  
 DB 531 SI-----DNNYIISQSDTLKLWRINHLVPLLLKKKEENDEQTKNEQSEQNEHNE 583

```

Qy 104 DVSKK---KONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVTL 148
|:|:| ||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 584 DYAKKTSKONDHANNOEDGSEKK-----KKKKEKNDKIKSKITLL 626

RESULT 28
ID Q9FJK9_ARATH PRELIMINARY; PRT; 470 AA.
AC Q9FJK9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gb|AAZ20218.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB015468; BAB10694.1; -; Genomic DNA.
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72E35AC54 CRC64;

Query Match 12.7%; Score 107.5; DB 2; Length 470;
Best Local Similarity 20.1%; Pred. No. 19;
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

Qy 19 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGFE-----GKK 70
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 82 RENRVTVTVQNSNGSEK-----YVDLARRIRYDE-EATQSQRIDHPNQK 129
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Qy 71 DAGVIVNLKSDTFIKPVFKKIEKKEBENKFTPDVSKKKON----- 111
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 130 NVGITERKAFENSPIEETSHRVDDNKRINQKNFTAAKSSSENAVSRVSGADHKRAEYVGK 189
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Qy 112 PQVNHSQLNE-----SHRKEDLQREHSQKSDSTKDVATVTLQKNISSKTTNPNK 164
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 190 PMENRDQVRQTESAEKSHRKENVTKSEKPRDQGVKTEAKDKDRNKKEKKEKTESINK 248
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

RESULT 29
Q640L5_MOUSE
ID Q640L5_MOUSE PRELIMINARY; PRT; 1455 AA.
AC Q640L5;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Sarcoma antigen NY-SAR-41
DE Name=Ccdcl18; Synonyms=4932411G06rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6; TISSUE=Head;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marchionni K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ugin T.B., Toshivuku S., Carninci P., Prange C.,
RP NUCLEOTIDE SEQUENCE.

```

```
QY 28 QNGKMSSTIVSEEDFILPVYKGL--EKG-----YQFGWISGPEGKQDAGYVNL-- 78
Db 19 QSGHPLKFSFWEKTHLQIYNRLREBGNKNVQEMSPSISGTEGKKEIQMISHLQL 78
QY 79 -----SKDFIKPVFKIEEKEENKPTFDVSKKKONPVNHSQNLNESHKEDLQREH 133
Db 79 QSGKHQDSVLSANDSNLKGANEASGNGKSDENVKKSD--EENAKKSD---EEN 134
QY 134 SOKSDSTKDVATVLDKNNISSKS---TTNN 161
Db 135 KDANSNTKDAESAEGEENPVSQENQMKTLNN 165

RESULT 31
Q5W5T1 TETPY
ID Q5W5T1_TETPY PRELIMINARY; PRT; 1015 AA.
AC Q5W5T1_
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Putative P-glycoprotein (Fragment).
GN Name=mdrl;
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymenidae; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GL;
RA Camares O., Denizau F., Bamad M.;
RT "Characterisation of MDR sequence homologue in Tetrahymena
pyriformis.";
RL Submitted (Oct-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AJ514918; CDS5936.2; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0016887; P:ATPase activity; IEA.
DR GO; GO:0042626; P:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR011527; ABC membrane 1.
DR InterPro; IPR001140; ABC TM transpt.
DR InterPro; IPR003439; ABC transp_like.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS0929; ABC TM1F; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 2.
DR PROSITE; PS00893; ABC TRANSPORTER_2; 2.
KW ATP-binding; Nucleotide-binding.
FT NON_TER 1
SQ SEQUENCE 1015 AA; 114219 MW; CB2E9AB73768A778 CRC64;

Query Match 12.6%; Score 106.5; DB 2; Length 1015;
Best Local Similarity 27.9%; Pred. No. 51;
Matches 43; Conservative 26; Mismatches 62; Indels 23; Gaps 6;

QY 12 TGEVSELPHRVTVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWISGPEGKD 71
Db 308 TGRITVIAHLT-TIRNAD-----ILVIDKGLVQCTFD--QLIDARGKFE 353
QY 72 AGYVNLKSDTPIKPVFKIEEKEENKPTFDVSKKKONPVNHSQNLNESH-RKEDLQR 130
Db 354 ALAKNQIQKQKDEERKCEQLQEDRK--DPEERMKSTVGKSLIMENHLTKEQIAA 410
QY 131 EEHSKSDSTKDVATVLDKNNISSKSTTNPNK 164
Db 411 EEEKEQAYFKE-----LDKNMWTFLFTWNRPER 439
```

```
RESULT 32
Q6BRW2 DEBHA
ID Q6BRW2_DEBHA PRELIMINARY; PRT; 616 AA.
AC Q6BRW2_
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA4458|IPF8464 Candida albicans IPF8464 unknown
DE function.
GN OrderedLocusName=DEHA0D146749;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Bayne E., Bleykasten C.,
RA Boissane A., Boyer J., Cattolico L., Confanieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR382136; CAG87226.1; -; Genomic DNA.
DR GO; GO:0016301; P:kinase activity; IEA.
DR InterPro; IPR000749; ATP-gua_Ptrans.
KW Complete proteome.
SQ SEQUENCE 616 AA; 72143 MW; 884009B2B8B6C3CP CRC64;

Query Match 12.5%; Score 106; DB 2; Length 616;
Best Local Similarity 30.7%; Pred. No. 32;
Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9;

QY 46 PVYGELEKGYQFGWISGPEGKKGAGYVNLKSDT-FIKPV----PKIEEKKKEENK 100
Db 87 PVLGRRAKPKIYLTDTIT-----KDIN-DINFSDSSEHKEPIETSKTKSKTKTKTK 140
QY 101 PTFDVSKKK-----DNPO--VNHSQNLN-----SHRKEDLQREHSQK-----SDSTKDV 144
Db 141 PDLDIGKLERRIVADNPDEIENHSESESEIKQRKEKQRKQDKQREKLKAKNQESNNDST 200
QY 145 ATVLDK-NNISSKSTTNPN 163
Db 201 TEQPEPLKNNINEKITSNEPS 220

RESULT 33
O96229 FLAF7
ID O96229_FLAF7 PRELIMINARY; PRT; 951 AA.
AC O96229_
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0680w.
GN Name=PFB0680w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;
```

```

RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalloom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perteau M.,
RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Barriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalloom S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Bartell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AB001410; RAC71925.2; -; Genomic_DNA.
DR PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4F CRC64;

Query Match 12.5%; Score 106; DB 2; Length 951;
Best Local Similarity 24.3%; Pred. No. 52;
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGEVSELKPHRVTV--VTQNGKEMSSSTVSEDFILPVYKGELEKGYQPDGWEI--SGF 66
DB 128 EEKINKNSDLRQNELNLQSK-----NEQDI-----NKNKGGKQ-----DISNSNA 170

QY 67 EGKQDAGVYINLSKDTFKIPVKFKIEEKEE-----ENKPTFD-----VSKKDNPP 112
DB 171 ENKQD-----VKEGVKLEIEKKEEKEEKSIDHKKVLEKSSDDHKVYENKSSDDH 218

QY 113 QVNHSQLNESHKREDLQR--EHSQKSDSTKQVTVATVLDKNIS--SKSTTNPNK 164
DB 219 KVEENKSSDDHKIEEKKVKEEHEEDEE-----DKKEKKSENKKNKDNK 262

RESULT 34
Q7RJ3C PLAYO PRELIMINARY; PRT; 605 AA.
AC Q7RJ3C;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein (Fragment).
GN Name=PY03340;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalloom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000712; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD REPEATS_1; 2.
DR PROSITE; PS00682; WD REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469PE8E0F7 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 674;
Best Local Similarity 23.5%; Pred. No. 39;
Matches 39; Conservative 28; Mismatches 64; Indels 35; Gaps 6;

```

```

RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000954; EAA22905.1; -; Genomic_DNA.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 605 AA; 69357 MW; 74035FFFF4D548483 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 605;
Best Local Similarity 24.6%; Pred. No. 34;
Matches 43; Conservative 37; Mismatches 58; Indels 37; Gaps 8;

QY 9 NKDTGEVSELKPHRVTVT--IQNGKEMSSSTVSE-----EDFILPVYKGELEKGYQPDG 60
DB 369 NQDVFDSLALKEYGDSIKGHEEYKEVYTSLSNDIQEHLIEDVWLPLY--ETYSNY----- 422

QY 61 WEISGFEGKQDAGVYINLSKDTFKIPVKFKIEEKEE--ENKPTFDVSK-----KKDNP 112
DB 423 -----ADYVENFDIMEYLKVELEANEPQKTRTLEDKPPINVLFTPLPDRTP 471

QY 113 QVNHSQLNESHKREDLQR--EHSQKSDSTKQVTVATVLDKNIS--SKSTTNPN 163
DB 472 VQONTQMDAPAKISDQGNDSQESLSQKIDSTYSGKDKISDIPSEKRSQSPD 526

RESULT 35
Q7RL7E PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RL7E;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalloom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000712; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD REPEATS_1; 2.
DR PROSITE; PS00682; WD REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469PE8E0F7 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 674;
Best Local Similarity 23.5%; Pred. No. 39;
Matches 39; Conservative 28; Mismatches 64; Indels 35; Gaps 6;

```

```
QY 10 KDTGEVSELKPHRVTTVTIONGKEMSTTVSE-----EDFILPVYKGELEKGVQPDGW 61
DB 481 KTRLLGHQKPVHRTQSPNGKFIASSSFDKSIIRWSGIDGTLYLAVFRGHVGPAYKI-AW 539
QY 62 EISGPEGKDGAGVNLNSKDTPIK-----PVFKKIRE-----KKEEENKPTFD 104
DB 540 SI-----DNYIISCSQDSTLKLWRINHLVPLLKKKEENGDPKSPKNDQKDKD 592
QY 105 V--SKKNQPNVHSQLNESHKEDLQREHSQKSDSTKDVATVYL 148
DB 593 XKDDQKQDKQNDNDHANNQEDGEKKKKKBEKKDKIKSKIKTL 638

RESULT 36
Q54GS1_DICDI
ID Q54GS1_DICDI PRELIMINARY; PRT; 1550 AA.
AC Q54GS1_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=DDB0188660;
OS Dictyostelium discoideum (Slime mold). Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
ON NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindeay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrrell B., Kuapa A.,
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFI01000221; EAL62484.1; --; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1550 AA; 174527 MW; D6D5DF547DC48E6 CRC64;

Query Match 12.4%; Score 105; DB 2; Length 1550;
Best Local Similarity 25.4%; Pred. No. 1e+02;
Matches 46; Conservative 24; Mismatches 59; Indels 52; Gaps 6;

QY 20 PHRVTTVTIONGKEMSTTVSEDFILPVYKGELEK--GYQPDGWEISGFEKGDAGYVI- 76
DB 1328 PHK-----LKKGNRGSSSNLSILPTLEDHLRTVSPKVDMPFTSPFDDTLDLKELIS 1383
QY 77 -----NLKDTFFIKPVKKTEKKEEENKPTFDVS----- 106
DB 1384 TDELNGNNNNNNNNNNPFPNKRNSNEVKVQPVQOEIETEKEKENIIGEDVEKETIS 1443
QY 107 -----KKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 160
DB 1444 IQEKIKKDDNDNNNIVEKEEIEIKODNIKEIKQDDSNKKEI-----NN--DNSTN 1494
QY 161 N 161
```

```
DB 1495 N 1495
RESULT 37
Q9NFV9_PLAFA
ID Q9NFV9_PLAFA PRELIMINARY; PRT; 329 AA.
AC Q9NFV9_
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=map3; falciparum.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okumu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252287; CAB85901.1; --; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON TER 1 1
FT NON_TER 329 329
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.3%; Score 104.5; DB 2; Length 329;
Best Local Similarity 24.3%; Pred. No. 21;
Matches 43; Conservative 30; Mismatches 63; Indels 41; Gaps 8;

QY 1 TTVEFPIIL-KDTGEVSELKPHRVTTVTIONGKEMSTTVSEDFILPVYKGELEKGYQFD 59
DB 144 TKTEYKAKNAYEAKNAYQKANQAVLKAKEASS-----YDYL----- 184
QY 60 GWEISGP--EGKKDAG-----YVNLKSDTFIKPVKKIKKEEENKPTFDVSKKDN 112
DB 185 GWEPGGVPEHKKEENMLSHLYSSKKNISKENDVDLDE--KEEAEETEKEEKE 243
QY 113 QVNHSQLNESHKEDLQREH-----SOKSDSTKDVATVLDKNNISKSTTN 161
DB 244 EETSEISEDEEEEEEKEEKEKQKQSNENNNDQKXDMEA-----QNLISKNNQN 295

RESULT 38
Q4YVY2_PLABE
ID Q4YVY2_PLABE PRELIMINARY; PRT; 736 AA.
AC Q4YVY2_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
```



```

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;
Qy 1 TTWKEFILNKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
Db ||| : : | | | | | : : | | : : | :
27 TTVKA-VVEETKVEEDEXP-----EGVEKSASFKEESDFFADLKESEKK----- 70
Qy 61 WEISGPEGKKGAGYVNLSDTPIKVPFKK---IEKKEEENKPTDYSKKKDNQVNHHS 117
Db : : : : | | | | | : : : | | | | | : : : | | : :
71 -ALSDLKSEEAIVDN-----TLLKTKKESPFMKKEEVEVVKPEAEVEKKKE--EAAEE 123
Qy 118 QLNESHKEDLQREEHHSQKSDSTKQVTVATVLDKNNISS 155
Db : : : : | | : : | | : : | : :
124 KVEEKKSEAVVTERAPKAETVEAVVTEIIPKEVTT 161

```

Search completed: April 24, 2006, 14:59:38  
Job time : 100.611 secs



**THIS PAGE BLANK (USP10)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 17.2438 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVKEFILNKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUMP62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	848	100.0	2140	2	F95074	serine proteinase,
2	845	99.6	2144	2	A97942	metalloproteinase
3	111	13.1	1038	2	JC5497	claustrin - chicke
4	110	13.0	558	2	T18467	hypothetical prote
5	106	12.5	665	2	B71609	hypothetical prote
6	104	12.3	540	2	D86432	hypothetical prote
7	103.5	12.2	325	2	T18283	hypothetical prote
8	101.5	12.0	312	2	G81339	probable membrane
9	100	11.8	622	2	A90570	lipoprotein impor
10	99.5	11.7	385	2	T20410	hypothetical prote
11	99	11.7	211	2	T25911	hypothetical prote
12	99	11.7	219	2	B72291	hypothetical prote
13	97.5	11.5	614	2	A84152	hypothetical prote
14	97.5	11.5	1345	2	S46817	hypothetical prote
15	97	11.4	988	2	T14188	hypothetical prote
16	96.5	11.4	456	2	T05612	hypothetical prote
17	96	11.3	535	2	T37189	hypothetical prote
18	96	11.3	2500	2	G71609	hypothetical prote
19	95.5	11.3	644	2	T47835	hypothetical prote
20	95	11.2	348	2	I37271	cyclin II - human
21	95	11.2	1397	2	T10466	DNA topoisomerase
22	94.5	11.1	867	2	T27136	hypothetical prote
23	94.5	11.1	871	2	T27135	hypothetical prote
24	94.5	11.1	3724	2	T18427	hypothetical prote
25	94	11.1	210	2	T28771	hypothetical prote
26	93.5	11.0	645	2	E89883	conserved hypotet
27	93.5	11.0	649	2	S42488	dnak-type molecula
28	93	11.0	629	2	G96542	hypothetical prote
29	93	11.0	2464	1	QRMSPI	microtubule-associ

ALIGNMENTS

RESULT 1

P95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: P95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hea  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf  
neon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, P.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso  
A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; PMID:21357209; PMID:11463916  
A:Accession: P95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: 1-2140 <KUR>  
A:References: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%	Score 848;	DB 2;	Length 2140;
Best Local Similarity	100.0%	Pred. No. 1.4e-53;		
Matches 164;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDPILPVYKGLKGYQFDG	60	
Db	1943	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDPILPVYKGLKGYQFDG	2002	
QY	61	WEISGFEKKDAGVYINLSKDTFKVPKLEKKERENKPTFDVSKKDNPNVNSQLN	120	
Db	2003	WEISGFEKKDAGVYINLSKDTFKVPKLEKKERENKPTFDVSKKDNPNVNSQLN	2062	
QY	121	ESHRKEDIQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK	164	
Db	2063	ESHRKEDIQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK	2106	

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; PMID:21429245; PMID:11544234

A:Accession: A97942  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUP>  
A:Cross-references: UNIPROT:Q8DQ7; UNIPARC:UPI0000E3490; GB:AE007317; PIDN:AAK99365.1;  
C:Genetics:  
A:Gene: prta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 845; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 2.3e-53;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 1947 TTVKEFILNDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2006

QY 61 WEISGPEGKADGYVINLSKDTFIKVPKKEEKEENKPTFDVSKKDNPNQVNSQLN 120  
DB 2007 WEISGPEGKADGYVINLSKDTFIKVPKKEEKEENKPTFDVSKKDNPNQVNSQLN 2066

QY 121 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
DB 2067 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2110

RESULT 3  
JC5497  
claustrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: JC5497; PC4334; S37561  
R:Burg, M.A.; Cole, G.U.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BURL>  
A:Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:G406318; PID  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Cross-references: UNIPARC:UPI000017BPF3; UNIPARC:UPI000017BPF4; UNIPARC:UPI000017BPF5  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 13.1%; Score 111; DB 2; Length 1038;  
Best Local Similarity 23.7%; Pred. No. 2;  
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

QY 3 VKEFILNDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTF 83  
DB 528 VQAKLKORTDSKESLKPAAKTTTKQDCQRNLKKHWSLSQVLOOLEKPKQLESKEKTP 587

QY 30 -----GKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTF 83  
DB 588 VKKERAKVPETKIIVAKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKES 639

QY 84 IKPVFK-KTEEKKEENKPTFDVSKKDNPNQVNSQLNESHKEDLORE-----EHSQKS 137  
DB 640 VKKEVKAKPEEKKEKPKKEVSKKEKPLI---KKEBKPKKEDIKKEVKKVKEKK 696

QY 138 DSTKDV 143  
DB 697 EAKVEV 702

RESULT 4  
TI8467  
hypothetical protein C0465C - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: TI8467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: TI8467  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1, 160/1  
A:Note: C0465C

Query Match 13.0%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 1.2;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 37 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEKGYK-----DAGYVINLSKDTFIKPV 87  
DB 60 ILGFEDDILYEYCISQLKQSKKE--KADGEBEDKYLNAKCLKINLTGPIGNKKSDFIEEL 117

QY 88 PKKI--BEKKE-----ENKPTFVSK-KKONPQVNSQLN-----SHRK 125  
DB 118 LELLNBEKKEEHTADTLNENK-TNCJIKVKVNEINENYVYENKQISNKDKHVSQHN 176

QY 126 E-----DLQREH-----SQKSDTK-----DVTATVLDKNNISSKSTTN 160  
DB 177 EHNINNVNLKKEKEYTDIQDRKHKRSLSQKSDSYKRPFNKRTKSTIER-SLSNRYDE 235

QY 161 NPNK 164  
DB 236 KTNK 239

RESULT 5  
B71609  
hypothetical protein PF0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; NID:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PF0680w

Query Match 12.5%; Score 106; DB 2; Length 665;  
Best Local Similarity 24.3%; Pred. No. 2.8;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGEVSELKPHRVV-TVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGF 66  
DB 127 EEKNKINKSLHRQNELNLQSGK-----NEQDI-----NKNEKGKQ-----DISNSNA 169

QY 67 EGKGDAGYVINLSKDTFIKVPKKEEKE-----EENKPTPD-----VSKKDNPN 112  
DB 170 ENKKD-----VKGEVKELEKKEEKEKSDDHKVEENKKSDDHKVEENKKSDDH 217

QY 113 QVNSQLNESHKEDLQR-EHHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164

Db 218 KVEENKSDHDKIEBVKVVEEHEDEE-----DKKEKSNKNDENK 261

RESULT 6

Db6432

hypothetical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: D86432

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huijzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, L.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali, C.A.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 <STO>

A:Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI00000ABF35; GB:AE005172; NID:g4587525; P1

C:Genetics:

A:Map position: 1

Query Match 12.3%; Score 104; DB 2; Length 540;

Best Local Similarity 25.3%; Pred. No. 3;

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

Qy 1 TTVKGFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60

Db 27 TTVKA-VVEETKVEDESKP-----EGVEKSASPKEESDFPADLKESEKK----- 70

Qy 61 WEISGFEKGDAGYVINLSKDTFIKPVFKK----IEEKKKEENKPTFDVSKKDNQVNVHS 117

Db 71 -ALSDLKSKLEAIVDN---TLATKKCKSSPMKEKKEEVVVKPEAVEKKE--EAAEE 123

Qy 118 QLVNESHRRKEDLQREHSHQSDSKDTVTATVLDKNISS 155

Db 124 KVEEKKSEAVVTEEPAPKAETVEAVVTEETIIPKEEVTT 161

RESULT 7

Ti8283

hypothetical protein G5 - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: Ti8283

R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh Genetics 148, 1117-1125, 1998

A:Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1

A:Reference number: Z14684; MUID:98198836; PMID:9539429

A:Accession: Ti8283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-325 <RI>

A:Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; P1

C:Genetics:

A:Introns: 85/1

Query Match 12.2%; Score 103.5; DB 2; Length 325;

Best Local Similarity 23.8%; Pred. No. 1.9;

Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

Qy 21 HRVTVTIQNGKMSSTIVSEEDFILPVYK-GEL--EKGYQFDGWEISGFEK----- 69

Db 57 HRTTISIKN--RPSVKKIGDEEKLFRISKNGELIVNLEDFDNFIK--EGKHLKSKMP 112

Qy 70 ---KDAGYV-----INLSKDTFIKPV-----FKK----- 90

Db 113 NHIKDSGYATNBEIEIFLESCITLCKEITACTGRNSYKRNIIINKLPEEEEEEEEEEE 172

Qy 91 --IEKKKEENKPTFDVSKKDNQVNVHSQLNESHKKE-----DLQREHSHQSDSK 141

Db 173 EEEBQEEVEKPTIIEBEEBETPAVSEEEKEEEEEEEBETPAVSEEEKEEEEEQEEDKEK 232

Qy 142 D 142

Db 233 D 233

RESULT 8

G81339

probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004

C:Accession: G81339

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Braham, D.; Chilli, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barr Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals b;

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: G81339

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-312 <PAR>

A:Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL139076; GB:AL111168; N

C:Genetics:

A:Gene: Cj0692c

Query Match 12.0%; Score 101.5; DB 2; Length 312;

Best Local Similarity 25.0%; Pred. No. 2.5;

Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

Qy 5 EFILNKDTGEVSELKPHRVTVTIQNGKMSSTI---VSEDFILPVYK-----GBLE 53

Db 5 DFIKERQNIROKMLKFSR---AINQKPLDLDLDEISSDILRRRPKKTKPNKPLEBLD 61

Qy 54 KGYQFDGWEISGFEKGDAGYVINL---SKDTFIKPVFKKIEEKKKEENKPT----FDYSK 107

Db 62 BEYESKTKTKSNLYLKED---LINVKLEEKOSLAKTIFSKMKERKKEENKTKKNFLFSR 118

Qy 108 KKDNP---QVNVHSQLNESHKEDLQREHSHQSDSKDTVDV--TATVLDKNISSK--STT 159

Db 119 KKANEIKNIQTKTQIQTKSNQATTTQKQKKELINSIEKIQTETKIQKPLIEKKLDVK 178

Qy 160 NPN 163

Db 179 NPN 182

RESULT 9

A90570

lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004

C:Accession: A90570

R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, P.; Moszer, I. Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: A90570

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-622 <KUR>

A:Cross-references: UNIPROT:Q980A1; UNIPARC:UPI00000C80A1; GB:AL445566; PID:g14089879;

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPU 4650

A:Genetic code: SGC3



Best Local Similarity 27.3%; Pred. No. 10;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;  
QY 16 SELKPHRVTTIQNGKMSSTIVSE-----BDFILPVYKGELEKGYQFDGW-----EISG 65  
DB 386 SBPKPE--TYTLQTAIQM--TPIVNEVSPQTRBEFL-----ARAHQLDGWADVSKVTY 435  
QY 66 FEGKKDAGVIVNLKSDTFIKPVKIEKK--EENKPTFDVSKKK-----DNPQVNHSL 119  
DB 436 FAGRTASQL-----GKIEEKLQKYNNTYTFDFKKEVNVQSPKSTGAL 482  
QY 120 NESHRKEDLQREHSQKSDTKDVTATVLD 149  
DB 483 GKVIGIVDLGRVHKHGEDLERLSKSQIE 512

## RESULT 14

S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C:Accession: S46817  
R:Pavello, T.  
submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46817  
A:Molecule type: DNA  
A:Residues: 1-1345 <FAV>  
A:Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:G500825; PID  
C:Genetics:  
A:Cross-references: SGD:S0001122  
A:Map position: 8R  
C:Superfamily: uncharacterized conserved protein  
C:Keywords: transmembrane protein

Query Match 11.5%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 25;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;  
QY 49 KGELEKGYQFDGWISGPEGKK-DAGYVIVNLKSDTFIKPVKIEKKKEENKPTFDVSK 107  
DB 1109 KGAIEKG-----SVGGQKVSDYMLSELRDI-----SRASKKPVKKVMK 1149  
QY 108 KCONPQVNHSLNESHRKEDLQREHSQKSDTKDVTATVLD--KNNISSKSTTN---NP 162  
DB 1150 SHDKRPHFSKVE-----QKSESRSKDDNKDILTHILDVQNNFSSEIFMKNKLSP 1201  
QY 163 NK 164  
DB 1202 QK 1203

## RESULT 15

T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z17931  
A:Accession: T14188  
A:Molecule type: DNA  
A:Residues: 1-988 <BEV>  
A:Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI00000A3618; EMBL:AL109819  
A:Experimental source: cultivar Columbia; BAC clone T28D5  
C:Genetics:  
A:Gene: ATSP:T28D5.30  
A:Map position: 4  
A:Introns: 162/3; 201/3; 416/3; 438/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C:Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 11.4%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 19;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;  
QY 29 NGKEMSSSTIVSEEDFILPVYKGELEKGY-----OPDGWEISGFEKGDAGYVIVNLKSDTF 83  
DB 383 NGRQNSNVQSVDEILSYTDTKVPVSGVGLNVSESDIVELVEDDVRSGAGLSPNVQRDN- 441  
QY 84 IKPVKKIIEKKK-----EENKPTFDVSKKKONPQVNHSLNESHRKEDLQRE 132  
DB 442 VEPVGDDVRSSGDSPNPSPANNVREGPATTFIMESDNPGRDNPVAPMEDHIRSEVQLSP 501  
QY 133 HSQKSDSTKDVTT--ATVLDKNNISSKSTTNPNK 164  
DB 502 HVL---GAKDVTDVSDPTDKRVGVNDVTDASDPT 532

## RESULT 16

T05612  
hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, B.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81P5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.4%; Score 96.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 8.8;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;  
QY 10 KDTGEVSLKPE-HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGPEG 68  
DB 245 KKKKKKPKPKPHVPSAPLVYVANERRAALREKNSKVVEVAK-----ITGEENKUNLSD 296  
QY 69 KKDAGY--VIVNLKSDTFIKPVKIEKKKEENKPTFDVSKKKONPQVNHSLNESHRKE 126  
DB 297 KKKAPYKRVAKNKETVYQ-AMEBYKRTKEE-----ALSQKE-----EELLLKLHKQE 345  
QY 127 DLQREHSQKSDTKDVTATVLDKNNISSKSTTN--PNK 164  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKKNNVDPNK 378

## RESULT 17

T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37189  
R:Leimbac, D.; Mixx, M.  
submitted to the EMBL Data Library, February 1996  
A:Description: The sequence of C. elegans cosmid C02H7.  
A:Reference number: Z20523  
A:Accession: T37189  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-535 <LEI>  
A:Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:U49945; PIDN:AAC47924.1  
A:Experimental source: strain Bristol N2; clone C02H7  
C:Genetics:  
A:Gene: CESP:C02H7.1  
A:Map position: X  
A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 96; DB 2; Length 535;  
Best Local Similarity 21.3%; Pred. No. 11;  
Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

QY 11 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKKG 70  
DB 77 DGSLKNVKAII---SGDAET-----NKMQLMGTNATSFNSRN 116

QY 71 DAGYVNLKOTPIKPVFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 130  
DB 117 GTG-----EKKKKKKKKEDKGDDEKST---TKRSSKKTHERKEKSEKKKSAAE 166

QY 131 EHSQKSDSTK-----DVTATVLDKNNISSKSTTNNPNK 164  
DB 167 KEKKKKSSSKERHKSSDRSSSEKSEKSEKKEKKEKSTTDEKPK 210

RESULT 18  
G71609  
hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: G71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: G71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2500 <GAP>  
A:Cross-references: UNIPROT:O96223; UNIPARC:UPI0000078301; GB:AE001408; GB:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0650w

Query Match 11.3%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 64;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 19 KPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKKGAGVNL 78  
DB 2173 KPKIT---ENKK-----NEGNEILKKYSIENEKKNYDKEQENCLDKDTQCNVT 2223

QY 79 -----SKDTFKPVFKIEKKEENKPTFDVSKKONPQVNHSQLN-----SHRKED 127  
DB 2224 KEKNLNDKKSPFNIKVKLEBEKSDDKRD---DKKNDTREKNLNDKKSPFNIKVK 2280

QY 128 LQREHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 164  
DB 2281 LEEEE---KSDDKRD-----DKKNDTREKNLNDK 2308

RESULT 19  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Starr, W.; Holland, R.; Weicheelgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NYA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3

A;Note: T209.90

Query Match 11.3%; Score 95.5; DB 2; Length 644;  
Best Local Similarity 23.4%; Pred. No. 15;  
Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

QY 28 QNGKEMSTIVSEB---DFILPVYKGELEKGYQDGEWISGFEKKGADGVNLNSKDTF 83  
DB 519 ENSKTEKTVADKKSVADFLKRIKKNSPQKG-----ETTSKNQKNDGNV----- 565

QY 84 IKPVFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDS----- 139  
DB 566 -----KXENDHOKSGDGNVKKENSKVKPRELSSSTGKKKVEVNNNNSSSKRKQ 615

QY 140 TKDVTATVLDKNNISSKSTTNNPNK 164  
DB 616 TKETAETVATGREGSKDKDKQPK 640

RESULT 20

I37271  
Cylclin II - human  
C:Species: Homo sapiens (man)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: I37271; S52774  
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.  
Exp. Cell Res. 218, 174-182, 1995  
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the  
A:Reference number: I37271; MUID:95255491; PMID:7737358  
A:Accession: I37271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-348 <HES>  
A:Cross-references: UNIPROT:Q14093; UNIPARC:UPI0000128C36; EMBL:Z46788; NID:g758586; PID

Query Match 11.2%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 8.4;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 49 KGELEKGYQDGEWISGFEKKGADGVNLNSKDTFIRKPVFKIEKKEENKPTF---DV 105  
DB 205 ESEGEKG---GTSEKSKKKKDS-----KKGKSAIBLQAVKADEKDEGKKDANKGDE 256

QY 106 SK--KONPQVNHSQLN-----ESHKEDLQREHSQKSDSTKD---VTATVLDKNNI 153  
DB 257 SKDAKKDAKEIKKGGKKKPSSTDSKDDVKE---SKDATKDAKKVAKDKTEKESA 313

QY 154 SSK 156  
DB 314 DSK 316

RESULT 21

T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falc  
C:Species: Plasmodium falciparum  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: T10466  
R:Cheesman, S.J.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Z17031  
A:Accession: T10466  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1397 <CHE>  
A:Cross-references: UNIPARC:UPI00000006D7; EMBL:X79345; NID:g994807  
C:Genetics:  
A:Gene: TopoII  
A:Map position: 14  
A:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.2%; Score 95; DB 2; Length 1397;



A;Molecule type: DNA  
A;Accession: 1-871 <WIL>  
A;Residues: 1-871  
A;Cross-references: UNIPROT:O62486; UNIPARC:UP10000079598; EMBL:Z99278; PIDN: CAB16492.1;  
A;Experimental source: clone Y53C12B  
C;Genetics:  
A;Gene: CESP:Y53C12B.3b  
A;Map position: 2  
A;Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-210 <NELL  
A; Cross-references: UN  
A; Experimental source:  
C; Genetics:  
A; Gene: CESP:B03H12.5  
A; Map position: 4  
A; Introns: 30/2; 201/3







[illegible][illegible]

C:Species: Melanoplus sanguinipes entomopoxvirus  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <NA>  
A:Cross-references: UNIPROT:O9YVL2; UNIPARC:UPI0000F2D3E; EMBL:AF063866; NID:g4049647;  
C:Genetics:  
A>Note: MSV230

Query Match 10.7%; Score 90.5; DB 2; Length 670;  
Best Local Similarity 22.5%; Pred. No. 37;  
Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;

QY 25 VTIQNGKMSSTI-----VSEEDFILPVYKGELEKGY-----QPDGWEISGF 66  
Db 33 VNFEKKQIISILLKFNDFKTEMCGVSEKVFQLINKSASEKSDVDSSIDESQNSDS 92  
QY 67 EGKGDAGYVINI-SKDTFIKPVFKKIE-EKKBEENKPTFDVSKKONPQVNHSQL-NESHR 124  
Db 93 DSDSDSGVNIIDSQNSDSKVINIKLENESQNSDSKVINIDESQNSDS-KVINIKLENESQN 151  
QY 125 KEDLQREHSQKSDSTKVDTATVLDKNNISSKSTNNPNK 164  
Db 152 SDSKVINIDESQNSDSKVINIDESQNSDSKVINIDESQNSDSK 191

RESULT 39  
S41552  
probable transcription factor SPT7 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YBR0739; protein YBR081C  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 05-Oct-2004  
C:Accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716  
R:Gansheroff, L.; Dollard, C.; Tan, P.; Winston, F.  
submitted to the EMBL Data Library, July 1993  
A:Reference number: S41552  
A:Accession: S41552  
A:Molecule type: DNA  
A:Residues: 1-1332 <GAM>  
A:Cross-references: UNIPROT:P35177; UNIPARC:UPI000005310C; EMBL:L22537; NID:g349189; PID:R.Steenma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45932  
A:Accession: S45946  
A:Molecule type: DNA  
A:Residues: 1-1332 <STB>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:Z35950; NID:g3536341; PIDN:CAA85026.1; PID:R.Andre, B.; Czapluch, C.; Hein, C.; Jaumiaux, J.C.; Urrestazu, A.; Visseers, S.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45893  
A:Accession: S45948  
A:Molecule type: DNA  
A:Residues: 1-835 <AND>  
A:Cross-references: UNIPARC:UPI000017A10D; EMBL:Z35950; MIPS:YBR081C  
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; David, I.B.  
Nucleic Acids Res. 20, 2603, 1992  
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.  
A:Reference number: S40800; MUID:92285152; PMID:1350857  
A:Accession: S40800  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 463-523 <HAY>  
A:Cross-references: UNIPARC:UPI0000168E4B; EMBL:M87651; NID:g172683; PIDN:AAA35087.1; PID:R.van der Aart, Q.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.  
Yeast 10, 959-964, 1994  
A:Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae.  
A:Reference number: S45462; MUID:95076715; PMID:7985423

A:Accession: S45478  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-624, 'LRGKKRKI', 633-1332 <VAN>  
A:Cross-references: UNIPARC:UPI000017A10E; EMBL:X76294  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1993  
R:Gansheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.  
Genetics 139, 523-536, 1995  
A:Title: The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for high-molecular mass  
A:Reference number: S54985; MUID:95229044; PMID:7713415  
A:Accession: S54985  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1332 <GAW>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:R.van der Aart, Q.J.M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S59702  
A:Accession: S59716  
A:Molecule type: DNA  
A:Residues: 1-1332 <VAM>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:X76294; NID:g974203; PIDN:CAAS3940.1; PID:A.Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SPT7  
A:Cross-references: SGD:S0000285; MIPS:YBR081C  
A:Map position: 2R  
C:Keywords: nucleus; transcription regulation  
F:466-521/Domain: bromodomain homology <BRO>

Query Match 10.7%; Score 90.5; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 80;  
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 25 VTIQNGKMSSTI VSEEDFILPVYKGELEKGYQPDGWEISGFEGKGDAGYVINLSKDTFI 84  
Db 549 ITIRNRADLEKEI---EDM-----EKDKDYELDEEEVAGSGRKG-----LNMGAHMLA 594  
QY 85 K---PVPKLTIEKKEENKPTFD-----VSKKKD-----NPQVNHSQLNESH 123  
Db 595 KENGKVSSEKSSKTVKDEAPTNDKLTSVIPEGEKEDKDTASSTVTVHENYVKNQEIENG 654  
QY 124 RKEDLQR--EEHSQKSDSTKVDTATVLD--KNNISSKSTTN 160  
Db 655 KNEQDMVESSKTESSKDAADAAKKTEDGLQDKTAEN 693

RESULT 40  
T28676  
rhoctry protein - Plasmodium yoelii (fragment)  
C:Species: plasmodium yoelii  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T28676; A45521  
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-2401 <SIN>  
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; PID:R.Keen, J.; Holder, A.A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoctry protein. Multiple clones  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KEE>  
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 10.7%; Score 90.5; DB 2; Length 2401;

Best Local Similarity 27.4%; Pred. No. 1.5e+02;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY	26	TIQNGKMSSTI-----VSEDFILPVYKGELEKGY--OPDGWEISGFEGKXDA	72
Db	924	TSKNHEEKISDIRKNSLKIIDFSESYINDI-KKELEKNVLESQNNNTDINQYLSKIEN	982
QY	73	GYVINLSKDTFIKPVFKKIEKKBEENKPTFDVSKKONPQVNHSQLNESHKEDLQREB	132
Db	983	IY--NILKLNKIKKIIDKVKVTDIEK-----NNKKINABELSNSEKIITQLKENSLLKE	1035
QY	133	HSQKSDSTKD-----VTATVLDKNNISS	155
Db	1036	COSKIKSTIDDNVSECIKNITNLKTYIVNEKNNT	1072

Search completed: April 24, 2006, 15:01:16  
Job time : 18.2438 secs



**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	119	14.0		746	6	US-10-793-626-652	Sequence 652, App
2	111.5	13.1		354	7	US-11-189-817-2	Sequence 2, Appli
3	110.5	13.0		651	7	US-11-128-660-1	Sequence 1, Appli
4	95	11.2		886	7	US-11-087-099-11456	Sequence 11456, A
5	93.5	11.0		501	6	US-10-485-517-381	Sequence 381, App
6	93.5	11.0		645	6	US-10-485-517-244	Sequence 244, App
7	92	10.8		299	6	US-10-793-626-1888	Sequence 1888, Ap
8	87	10.3		140	7	US-11-096-568A-4771	Sequence 4771, Ap
9	86	10.1		443	7	US-11-188-298-1015	Sequence 1015, Ap
10	86	10.1		700	7	US-11-196-475-74	Sequence 74, Appl
11	85.5	10.1		734	7	US-11-072-512-2272	Sequence 2272, Ap
12	85	10.0		472	6	US-10-793-626-658	Sequence 658, App
13	85	10.0		493	7	US-11-096-568A-3070	Sequence 3070, Ap
14	85	10.0		493	7	US-11-096-568A-3071	Sequence 3071, Ap
15	85	10.0		510	7	US-11-096-568A-3069	Sequence 3069, Ap
16	85	10.0		636	6	US-10-485-517-170	Sequence 170, App
17	85	10.0		700	7	US-11-196-475-66	Sequence 66, Appl
18	85	10.0		708	7	US-11-196-475-76	Sequence 76, Appl
19	85	10.0		720	6	US-10-793-626-2058	Sequence 2058, Ap
20	85	10.0		785	6	US-10-793-626-264	Sequence 264, App
21	84.5	10.0		1694	7	US-11-052-554A-83	Sequence 83, Appl
22	83.5	9.8		248	6	US-10-793-626-464	Sequence 464, App
23	83.5	9.8		635	6	US-10-821-234-1573	Sequence 1573, Ap
24	83.5	9.8		732	6	US-10-469-469-318	Sequence 318, App
25	83.5	9.8		732	7	US-11-187-230-21	Sequence 21, Appl



QY 57 QP-----DGEISGPEGK-----D 71  
Db 238 MVMETNDYKDFWVGQRTISKDAQNRTIIPFYVEGKTLYDAIVKVHVHTIDYD 297  
QY 72 AGVINL-SKDTFIKVPFKKIBKKKEENKPTFDV-----SKKDNPOVNHSQLNESHK 125  
Db 298 GQYHVRVDKEAFTKANTDKSNKQODNSAKKEATPATPSKPTSPVEKESQKQDSQKD 357  
QY 126 EDLQ-----REHSQKSDSTQV--ATVLDKNNISSKSTNNPNK 164  
Db 358 DNKQLPSVEKENDASSESGDKTPTKPTKGEVSSSTI--PTK 399

RESULT 6  
US-10-485-517-244  
; Sequence 244, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 244  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-244

Query Match 11.0%; Score 93.5; DB 6; Length 645;  
Best Local Similarity 21.9%; Pred. No. 3.4;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TVVKEFILNDTGR-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56  
Db 327 SAITEFQNVQPTNERMTDLQTKYVYVESVENNESMDTFVKH-----PIKTGLNGKKY 381  
QY 57 QP-----DGEISGPEGK-----D 71  
Db 382 MVMETNDYKDFWVGQRTISKDAQNRTIIPFYVEGKTLYDAIVKVHVHTIDYD 441  
QY 72 AGVINL-SKDTFIKVPFKKIBKKKEENKPTFDV-----SKKDNPOVNHSQLNESHK 125  
Db 442 GQYHVRVDKEAFTKANTDKSNKQODNSAKKEATPATPSKPTSPVEKESQKQDSQKD 501  
QY 126 EDLQ-----REHSQKSDSTQV--ATVLDKNNISSKSTNNPNK 164  
Db 502 DNKQLPSVEKENDASSESGDKTPTKPTKGEVSSSTI--PTK 543

RESULT 7  
US-10-793-626-1888  
; Sequence 1888, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1888

Query Match 10.8%; Score 92; DB 6; Length 299;  
Best Local Similarity 24.4%; Pred. No. 1.8;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;

QY 9 NKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFI-----LPVYKGELEK--YQPDGWEI 63  
Db 33 NKDT-EKSDKKYHRIISLIPSNTEILYRLIGIGEDIVGVSTVDDYPKDVKGKQKQFDAMNL 91  
QY 64 SGPE-----GKKDAGVYVNLSDTTPKV-----FKKIEKK 95  
Db 92 NKBELIKAKPDLILAHESQKNSAGKVLKSLDKGVVYVVKDAQSIDETDTPKSIQULT 151  
QY 96 EENKPTPDVSKKDNPOVNHSQLNESHKEDLQREHHSQ 135  
Db 152 DREKQAKELVDTEKHNVKEIINSVPKHKKQEVPMVEVSSK 191

RESULT 8  
US-11-096-568A-4771  
; Sequence 4771, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 4771  
; LENGTH: 140  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(140)  
; OTHER INFORMATION: Ceres Seq. ID no. 14304111  
US-11-096-568A-4771

Query Match 10.3%; Score 87; DB 7; Length 140;  
Best Local Similarity 25.9%; Pred. No. 1.8;  
Matches 36; Conservative 28; Mismatches 43; Indels 32; Gaps 7;

QY 30 GKEMSSIT-----VSEEDFILPVY--KGELEKGYQPDGWEISGPEGKQDAGYVNLSDT 82  
Db 8 GKQGSASAREDHGSGEEDKIPAYRRGRPKQPKMKD-----FEEDEDEB----- 52  
QY 83 FIKVPVKIEKKEENKPTPDVSKKDNPO-----VNHSQLNESHKEDLQREHHSQKSD 138  
Db 53 --BELVEKEEEEEEDSVT---SKKEENKRWINGSYTDANEKENGLOKSKSRDQSS 107  
QY 139 STKDVATVLDKNNISSKS 157  
Db 108 STKS-TSTGFRQNGSRKKS 125

RESULT 9  
US-11-188-298-1015  
; Sequence 1015, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B

```
/ CURRENT APPLICATION NUMBER: US/11/188,298
/ CURRENT FILING DATE: 2005-07-22
/ PRIOR APPLICATION NUMBER: 60/592,978
/ PRIOR FILING DATE: 2004-07-31
/ NUMBER OF SEQ ID NOS: 22569
/ SEQ ID NO 1015
/ TYPE: PRT
/ ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match          10.1%; Score 86; DB 7; Length 443;
Best Local Similarity 17.8%; Pred. No. 9;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 48 YKELEKGYQPDGWEISGEGKKDAGYVNLKDTFI-----84
Db 42 YEEAVEDGVSPGSSIPGEGIEDSLIFKADPSTYAEIPWEGIGRVGYIYKGDPEYQA 101
QY 85 --KPVFKIEEKKEEN-----KPTPDVSKKD-----110
Db 102 DPGILKRVLEKEKLEKKAHIGPEPFYFKNGTWELHIPDSGGYFDLVGLDKARIR 161
QY 111 -----NPQVNSQLNESHKEDLQREHSQKSD---STKDVATATVLD 149
Db 162 REIALYMPYLGLKPEVLHHEVGKAQHEIDFRYDEALRTADNIVSPKVVYKVAE 215

RESULT 10
US-11-196-475-74
/ Sequence 74, Application US/11/196,475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Constructs of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ PRIOR FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ PRIOR FILING DATE: 2001-08-07
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: Fastseq for Windows Version 4.0
/ SEQ ID NO 74
/ LENGTH: 700
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match          10.1%; Score 86; DB 7; Length 700;
Best Local Similarity 23.2%; Pred. No. 16;
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;

QY 23 VTTTONGKEMSTIVSEEDFILPVYKGELEKGYFD-----GWEISGEGKK 70
Db 258 ITETIENLADQEKATDEE-----HKKEIES--QVDAKKQKKEELDKAIDLAKAQQKL 309
QY 71 D-AGYVNLKDTFTFKPVFKIEEKKEENK-PDVSCKKNPQVN-----HSQLN 120
Db 310 DPAENLDIQDTRVREKQINETNKNLPKPGDVSPKVDKQLQKESLEDLQEQLK 369
QY 121 ES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTTNNPNK 164

/ CURRENT APPLICATION NUMBER: US/11/188,298
/ CURRENT FILING DATE: 2005-07-22
/ PRIOR APPLICATION NUMBER: 60/592,978
/ PRIOR FILING DATE: 2004-07-31
/ NUMBER OF SEQ ID NOS: 22569
/ SEQ ID NO 1015
/ TYPE: PRT
/ ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Db 370 ETGDENQRIEIKQIEIKKSKDEKLLKSKODKASKGKALDLDR-ELNSKASSKEKSK 425

RESULT 11
US-11-072-512-2272
/ Sequence 2272, Application US/11/072,512
/ Publication No. US20060029945A1
/ GENERAL INFORMATION:
/ APPLICANT: ISOGAI, TAKAO
/ APPLICANT: SUGIYAMA, TOMOYASU
/ APPLICANT: OTSUKI, TETSUJI
/ APPLICANT: WAKAMATSU, AI
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUUKO
/ APPLICANT: HIO, YURI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHICO
/ APPLICANT: YOSHIKAWA, TSUTOMU
/ APPLICANT: OTSUKA, MOTOTYUKI
/ APPLICANT: NAGAHARI, KENJI
/ APPLICANT: MASUHO, YASUHIKO
/ TITLE OF INVENTION: Novel full length cDNA
/ FILE REFERENCE: 084335-0191
/ CURRENT APPLICATION NUMBER: US/11/072,512
/ CURRENT FILING DATE: 2005-03-07
/ PRIOR APPLICATION NUMBER: US 60/350,978
/ PRIOR FILING DATE: 2002-01-25
/ PRIOR APPLICATION NUMBER: JP 2001-379298
/ PRIOR FILING DATE: 2001-11-05
/ NUMBER OF SEQ ID NOS: 4096
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 2272
/ LENGTH: 734
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match          10.1%; Score 85.5; DB 7; Length 734;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 39; Conservative 42; Mismatches 58; Indels 41; Gaps 11;

QY 10 KDTG-EVSELPKPRVTV-----TIQNGKEMSTIVSE--DFILPVYK 49
Db 168 EDTGIYVSEVDPNSIAAKDGRIRREGDRILQINGEDVQNRREAVALLSNDECKRIVLLVAR 227
QY 50 GELEKGYQPDGWEISGEGKKDAGYVNLKDTF-----IKPVFKIEE-KKEEENK 101
Db 228 PEIQLD---EGW-----LEDERNE-FLEELNLEMEHEEHEAMQPTANEVEQPKQEEEG 279
QY 102 TFDVSKKNPQVNHSQLNESHK-KEDLQREHSQKSDSTKDVATATVL-DKNNISSKSTT 159
Db 280 TTDATSSN---NHEKDSGVGRTDESURNDESSEQENNAEDPNSTSLKSRDLQSQSDT 336

RESULT 12
US-10-793-626-658
/ Sequence 658, Application US/10/793,626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match          10.0%; Score 85; DB 6; Length 472;
Best Local Similarity 32.6%; Pred. No. 12;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 91 IEKKEENKPTFDVSKKDNFQVNHSQLNESHKEDLQ-----REEHSQKSDS---- 139
Db 3 MEENKQPNKE--NMSKDDNA-----TLNDSHRNEDLELFRNKNARQRRRRIDNQSK 56

QY 140 TKDVTAT-----VLDKNNISSKSTTNNPNK 164
Db 57 EKDATSTQSLQETKPMDFLDNHS--HNQNK 86

RESULT 13
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

Query Match          10.0%; Score 85; DB 7; Length 493;
Best Local Similarity 26.3%; Pred. No. 12;
Matches 54; Conservative 23; Mismatches 74; Indels 54; Gaps 11;

QY 1 TTVKEFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTIVSEE---- 41
Db 42 TTKKKKIIKRVPKKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTIVTEKKPG 101

QY 42 -----DFILPVYKGELEKGYQFDGWEISGFGKKDAGYVI-----NLSKD 81
Db 102 KVPVKKIKITPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDDTQSTGKQTANAD 158

QY 82 TPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREEHHSQKSDS 139
Db 159 TTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTTKSDKDDKKEER-----V 208

QY 140 TKDVTATVLDKNNISSKSTTNNPNK 164
Db 209 TGEKSGAKTDKLSKSDKDVTVNKGK 233

RESULT 15
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match          10.0%; Score 85; DB 7; Length 510;
Best Local Similarity 26.3%; Pred. No. 13;
Matches 54; Conservative 23; Mismatches 74; Indels 54; Gaps 11;

QY 1 TTVKEFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTIVSEE---- 41
Db 59 TTKKKKIIKRVPKKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTIVTEKKPG 118

QY 42 -----DFILPVYKGELEKGYQFDGWEISGFGKKDAGYVI-----NLSKD 81
Db 119 KVPVKKIKITPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDDTQSTGKQTANAD 175

QY 82 TPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREEHHSQKSDS 139
Db 176 TTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTTKSDKDDKKEER-----V 225
```

```

; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3071
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match          10.0%; Score 85; DB 7; Length 493;
Best Local Similarity 26.3%; Pred. No. 12;
Matches 54; Conservative 23; Mismatches 74; Indels 54; Gaps 11;

QY 1 TTVKEFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTIVSEE---- 41
Db 42 TTKKKKIIKRVPKKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTIVTEKKPG 101

QY 42 -----DFILPVYKGELEKGYQFDGWEISGFGKKDAGYVI-----NLSKD 81
Db 102 KVPVKKIKITPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDDTQSTGKQTANAD 158

QY 82 TPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREEHHSQKSDS 139
Db 159 TTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTTKSDKDDKKEER-----V 208

QY 140 TKDVTATVLDKNNISSKSTTNNPNK 164
Db 209 TGEKSGAKTDKLSKSDKDVTVNKGK 233

RESULT 15
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match          10.0%; Score 85; DB 7; Length 510;
Best Local Similarity 26.3%; Pred. No. 13;
Matches 54; Conservative 23; Mismatches 74; Indels 54; Gaps 11;

QY 1 TTVKEFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTIVSEE---- 41
Db 59 TTKKKKIIKRVPKKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTIVTEKKPG 118

QY 42 -----DFILPVYKGELEKGYQFDGWEISGFGKKDAGYVI-----NLSKD 81
Db 119 KVPVKKIKITPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDDTQSTGKQTANAD 175

QY 82 TPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREEHHSQKSDS 139
Db 176 TTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTTKSDKDDKKEER-----V 225
```

4



Db 731 D-----DQTGNAIHTGEDAEMSEVFPENDDRIDIRQDFMDRVNVEDIESASDN 779

RESULT 21

US-11-052-554A-83

Sequence 83, Application US/11052554A

Publication No. US20050288866A1

GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.

TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

FILE REFERENCE: 30853/40359A

CURRENT APPLICATION NUMBER: US/11/052,554A

CURRENT FILING DATE: 2005-02-07

PRIOR APPLICATION NUMBER: US 60/589,227

PRIOR FILING DATE: 2004-07-20

PRIOR APPLICATION NUMBER: IN 173/DEL/2004

PRIOR FILING DATE: 2004-02-06

NUMBER OF SEQ ID NOS: 763

SOFTWARE: Patent in version 3.3

SEQ ID NO 83

LENGTH: 1694

TYPE: PRT

ORGANISM: Haemophilus influenzae Rd

US-11-052-554A-83

Query Match 10.0%; Score 84.5; DB 7; Length 1694;

Best Local Similarity 26.1%; Pred. No. 63;

Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFFIKVPVKKIEKKKEENKPTFDVSKKONQVNHSQLNESHRKEDLQREEHS- 134

Db 1288 INTGSAITATETAEKSKDPQTETAASTEDASQHKANTVADNSVANNSESSDPKSRRRSI 1347

QY 135 --OKSDSTKDVATVLDKNNISSKSTNNPNK 164

Db 1348 SQPQTSABETTAASTDETTIADNSKRKPNR 1379

RESULT 22

US-10-793-626-464

Sequence 464, Application US/10793626

Publication No. US20050255478A1

GENERAL INFORMATION:

APPLICANT: KIMMERLY, WILLIAM JOHN

TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

FILE REFERENCE: P03480US

CURRENT APPLICATION NUMBER: US/10/793,626

CURRENT FILING DATE: 2004-03-04

PRIOR APPLICATION NUMBER: 60/164,258

PRIOR FILING DATE: 1999-11-09

NUMBER OF SEQ ID NOS: 4472

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 464

LENGTH: 248

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: amino acid sequence

US-10-793-626-464

Query Match 9.8%; Score 83.5; DB 6; Length 248;

Best Local Similarity 24.8%; Pred. No. 7.1;

Matches 41; Conservative 18; Mismatches 65; Indels 41; Gaps 6;

QY 8 LNKDGTGEVSELKPHRVTVTTIQNGKMS-----STIVSEEDFILFVYKGELEKG 55

Db 50 LTADTTEVHKLEDYLK-TVEN-KEKSIYDLKRPVDLCNRSIKONEDIL--DYTKLPEKN 105

QY 56 YQPDGWEISFGEGKQAGYVINLSKDTFIKVPFKIEKKKEENKPTFDVSKKONQVNH 115

Db 106 RTEVESDINKAQNKEA-----SOLASKLENNQQLKDTAKK----- 142

```
QY 116 HSQLNESHKEDLQREHSQKSDSTKQVATATVLDKNINSSKSTTN 160
Db 143 --YLNSSNDSAKAIAKHISPLDKQITDKINTINSDNHVDN 195

RESULT 23
US-10-821-234-1573
; Sequence 1573, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1573
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1573

Query Match 9.8%; Score 83.5; DB 6; Length 635;
Best Local Similarity 23.6%; Pred. No. 23;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;

QY 6 FILNKDTGEVSELKPHRVTVTTONGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
Db 73 FTVRTDTGE-----PMGRGTVILHLKEDQTEYLEERRI-----KEIVKXHSQFIGYPITL 123

QY 66 P-EGKKDAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSKKDNPNQVNHSQLNESH 124
Db 124 FVEKERDK-----EVSDD-----EABEKEDKEEKEEKESEDKPEI----- 161

QY 125 KEDLQREHSQKSDSTKQVATVLDKN-----NISKSTTNPN 163
Db 162 -EDVGSDEEEKKDGDKKKKKIKEYIDQBELNKTPIWTRNPD 205

RESULT 24
US-10-469-469-318
; Sequence 318, Application US/10469469
; Publication No. US20060079493A1
; GENERAL INFORMATION:
; APPLICANT: FRITZ, LAWRENCE C.
; APPLICANT: BURROWS, FRANCIS J.
; TITLE OF INVENTION: METHODS FOR TREATING GENETICALLY-DEFINED PROLIFERATIVE
; DISORDERS WITH HSP90 INHIBITORS
; FILE REFERENCE: CON-0010-USN
; CURRENT APPLICATION NUMBER: US/10/469,469
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: PCT/US02/06518
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/272,751
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 318
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-469-469-318

Query Match 9.8%; Score 83.5; DB 6; Length 732;
Best Local Similarity 23.6%; Pred. No. 27;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;
```

```
QY 6 FILNKDTGEVSELKPHRVTVTTONGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
Db 170 FTVRTDTGE-----PMGRGTVILHLKEDQTEYLEERRI-----KEIVKXHSQFIGYPITL 220

QY 66 P-EGKKDAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSKKDNPNQVNHSQLNESH 124
Db 221 FVEKERDK-----EVSDD-----EABEKEDKEEKEEKESEDKPEI----- 258

QY 125 KEDLQREHSQKSDSTKQVATVLDKN-----NISKSTTNPN 163
Db 259 -EDVGSDEEEKKDGDKKKKKIKEYIDQBELNKTPIWTRNPD 302

RESULT 25
US-11-187-230-21
; Sequence 21, Application US/11187230
; Publication No. US20060035837A1
; GENERAL INFORMATION:
; APPLICANT: Altieri, Dario C.
; APPLICANT: Plescia, Janet
; APPLICANT: Saiz, Whitney
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT HSP90
; FILE REFERENCE: 07917-271001
; CURRENT APPLICATION NUMBER: US/11/187,230
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: US 60/590,584
; PRIOR FILING DATE: 2004-07-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-187-230-21

Query Match 9.8%; Score 83.5; DB 7; Length 732;
Best Local Similarity 23.6%; Pred. No. 27;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;

QY 6 FILNKDTGEVSELKPHRVTVTTONGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
Db 170 FTVRTDTGE-----PMGRGTVILHLKEDQTEYLEERRI-----KEIVKXHSQFIGYPITL 220

QY 66 P-EGKKDAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSKKDNPNQVNHSQLNESH 124
Db 221 FVEKERDK-----EVSDD-----EABEKEDKEEKEEKESEDKPEI----- 258

QY 125 KEDLQREHSQKSDSTKQVATVLDKN-----NISKSTTNPN 163
Db 259 -EDVGSDEEEKKDGDKKKKKIKEYIDQBELNKTPIWTRNPD 302

RESULT 26
US-11-096-568A-28315
; Sequence 28315, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28315
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1036)
```

```
; OTHER INFORMATION: Ceres Seq. ID no. 2712010
US-11-096-568A-28315

Query Match
Best Local Similarity 9.8%; Score 83.5; DB 7; Length 1036;
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVEPILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51
Db 492 TSSNKKIVKKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 548
QY 52 LEKGYPDQWEISGPEGKK--DAGYVINLSKDTFIKVPFKIEBKKEEENKPTDFVSKKK 109
Db 549 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 600
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVTLDKNNISSK 156
Db 601 GSPDTKKKGASSSSKKOTKTGDKKAEKKNNSETMSEGGKKIDRNNITDEK 650

RESULT 27
US-11-096-568A-28314
; Sequence 28314, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

Query Match
Best Local Similarity 9.8%; Score 83.5; DB 7; Length 1070;
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVEPILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51
Db 526 TSSNKKIVKKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 582
QY 52 LEKGYPDQWEISGPEGKK--DAGYVINLSKDTFIKVPFKIEBKKEEENKPTDFVSKKK 109
Db 583 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 634
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVTLDKNNISSK 156
Db 635 GSPDTKKKGASSSSKKOTKTGDKKAEKKNNSETMSEGGKKIDRNNITDEK 684

RESULT 28
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
```

```
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313

Query Match
Best Local Similarity 9.8%; Score 83.5; DB 7; Length 1276;
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVEPILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51
Db 732 TSSNKKIVKKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 788
QY 52 LEKGYPDQWEISGPEGKK--DAGYVINLSKDTFIKVPFKIEBKKEEENKPTDFVSKKK 109
Db 789 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 840
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVTLDKNNISSK 156
Db 841 GSPDTKKKGASSSSKKOTKTGDKKAEKKNNSETMSEGGKKIDRNNITDEK 890

RESULT 29
US-11-096-568A-17853
; Sequence 17853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17853
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(305)
; OTHER INFORMATION: Ceres Seq. ID no. 12361233
US-11-096-568A-17853

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 305;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGKXDAGYVINLS 79
Db 40 PDLVDCSLNCGDAGSSKKKAEKSSFRPVAK-----ETPSLEDSEKKKTKQAS 87
QY 80 KDTFIKVPFKIEBKKEEENKPTDFVSKKKNDPNFQVNHSQLNESHKEDL 128
Db 88 NQHSVK---KDIIEESNESVVKPQVGVSTPSYGFSGFKCDERAEKREFYFYSKLEEKIHAQEL 144
QY 129 QREHSQKSDSTKDVATVTLDKN-----TFDVSKKKNDPNFQVNHSQLNESHKEDL 151
Db 145 EKSNIQAKSKETBEAELOKLRKSLNFKATPMPSFYKPEPPPKVELKKIPTTTRASPGLGR 204
QY 152 --NISSKSTTNPN 163
Db 205 SKNTSSGGTEGNPN 218

RESULT 30
US-11-096-568A-17852
; Sequence 17852, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17851
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(327)
; OTHER INFORMATION: Ceres Seq. ID no. 12361232
US-11-096-568A-17851

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 327;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGVYNLS 79
Db 62 PLVDVCSLNGDAGSKKAEKSSRPVAK-----ETPSLEDSNEKKKTQKAS 109
QY 80 KDTFIKVPFKKIEKKERENK-----TFDVSKKKDNPOVNHSQLNESHKEDL 128
Db 110 NQHSVK---KDIEESNESVKPQRVGSTPSYGFSPKCDERAEKRREFYSKLEEKIHAQEL 166
QY 129 QREHSQKSDSKQVATVLDKN-----TQVSKKKDNPOVNHSQLNESHKEDL 151
Db 167 EKSNLQAKSKETEAEELKWLKSLNFKATPMPSPFYKEPPPKVELKKIPTTRARSPKLG 226
QY 152 --NISKSTNNPN 163
Db 227 SKNTSSGGTEGNPN 240

RESULT 31
US-11-096-568A-17851
; Sequence 17851, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17851
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12361231
US-11-096-568A-17851

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 386;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGVYNLS 79
Db 121 PLVDVCSLNGDAGSKKAEKSSRPVAK-----ETPSLEDSNEKKKTQKAS 168
QY 80 KDTFIKVPFKKIEKKERENK-----TFDVSKKKDNPOVNHSQLNESHKEDL 128
Db 169 NQHSVK---KDIEESNESVKPQRVGSTPSYGFSPKCDERAEKRREFYSKLEEKIHAQEL 225
QY 129 QREHSQKSDSKQVATVLDKN-----TQVSKKKDNPOVNHSQLNESHKEDL 151
```

```
Db 226 EKSNLQAKSKETEAEELKWLKSLN?KATPMPSPFYKEPPPKVELKKIPTTRARSPKLG 285
QY 152 --NISKSTNNPN 163
Db 286 SKNTSSGGTEGNPN 299

RESULT 32
US-11-087-099-5657
; Sequence 5657, Application US/11087093
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5657
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Bacillus anthracis str. Ames
US-11-087-099-5657

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 1070;
Matches 37; Conservative 24; Mismatches 66; Indels 27; Gaps 6;

QY 12 TGEVSELKP-----HRVTVTIQNGKEMSTIVSEEDFILPVY-----KGELEKGY 56
Db 411 TNEIRDTRPQELGKRIKIDAQKQKVLDEAYMNEEVKIPYVDVNGTALQNIWKSEGS 470
QY 57 QFDG---WEISG---FEGKKDAGVYNLSKDTFIKVPFKKIEKKERENKPTFDVSKK- 109
Db 471 ITNGVIKWSLGEKWEYFKMDAGE---SKIRFGQRVIONIVKRESSNVIQDKLQRY 526
QY 110 -DNPQVNHSQLNESHKEDLQREHSQKSDSKD 142
Db 527 MNKHNFERNKVNPTITKEDLLTVKALKITDGKCE 560

RESULT 33
US-11-096-568A-4567
; Sequence 4567, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4567
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(339)
; OTHER INFORMATION: Ceres Seq. ID no. 15219423
US-11-096-568A-4567

Query Match
Best Local Similarity 9.7%; Score 82; DB 7; Length 339;
Matches 43; Conservative 32; Mismatches 56; Indels 36; Gaps 9;

QY 9 NKOTGEVSELKPHRVTTVTIQNGKEMSTIVSE-----EDFILPVYK---GELEKGY 55
Db 114 NGKTGDVE--RPRRV-YDRRSQGRSDNVKREGGGRGNWGTPEDDIQPVTEPTTEVEKS 170
QY 56 --YQPDGWEISGFEGKKDAGVIN--LSKDTFIKVPFKKIEKK-----EENK---P 101
```

Db 171 PVAKEGGEDATTDAKGEAPEVEQEPEDKEMTLEBYEKLEKKALQATKVEERKVDTK 230  
QY 102 TFDVSKKDNQPNVHNSQL-----NESHKEDLQREHSQKSDSTKD 142  
Db 231 VPESMQLSNKKTNDSEIFIKLGDKDKKDAAEKEERAKKSLSINE 277

## RESULT 34

US-11-188-298-15964  
; Sequence 15964, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; PRIOR FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 15964  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus DSM 3638  
US-11-188-298-15964

Query Match 9.7%; Score 82; DB 7; Length 439;  
Best Local Similarity 18.1%; Pred. No. 19;  
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;  
QY 55 GYQPDGWSISGPEGKKGAGYVNLKDTFI-----KP-----VFK 89  
Db 45 GISPDGSSVPGQIGEDSLVFKADPDITVEVPMDNVARVGYFKDKNPKYGADPRGILK 104  
QY 90 KIEEKKEREN-----KPTDVSKKDN-----111  
Db 105 RALSELEKEGYKAYIGPPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIASYM 164  
QY 112 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATATVLDKNNI 153  
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKATADNIVSPKYITKAVAEHGL 215

## RESULT 35

US-11-188-298-16606  
; Sequence 16606, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; CURRENT FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 16606  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Pyrococcus woesei  
US-11-188-298-16606

Query Match 9.7%; Score 82; DB 7; Length 439;  
Best Local Similarity 18.1%; Pred. No. 19;  
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;  
QY 55 GYQPDGWSISGPEGKKGAGYVNLKDTFI-----KP-----VFK 89  
Db 45 GISPDGSSVPGQIGEDSLVFKADPDITVEVPMDNVARVGYFKDKNPKYGADPRGILK 104  
QY 90 KIEEKKEREN-----KPTDVSKKDN-----111

Db 105 RALSELEKEGYKAYIGPPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIASYM 164  
QY 112 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATATVLDKNNI 153  
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKATADNIVSPKYITKAVAEHGL 215

## RESULT 36

US-10-793-626-2482  
; Sequence 2482, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2482  
; LENGTH: 568  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-10-793-626-2482

Query Match 9.7%; Score 82; DB 6; Length 568;  
Best Local Similarity 23.2%; Pred. No. 26;  
Matches 38; Conservative 28; Mismatches 54; Indels 44; Gaps 7;  
QY 8 LNKDTGEVSELKPHRVTVTIQ-----NGKEMSTIVSEEDFILPVYKGELEKGYQ 57  
Db 199 LNEATTQLQEKVQKNOQPLREDMSFLTRFNKHAVNEKILNQETDVVPSLISEL---YQ 255  
QY 58 PDGWEISGFEKGKDGAGYVNLKDTFIKPVPKIEEKKKEENKPTFDVSKKKNPQVNH 117  
Db 256 TQ-TSISN-----TIVL-----TFSDEVIKALNKKIENESTPLFEEA-----VNHV 295  
QY 118 QLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISKSTNN 161  
Db 296 QVNEISSLSDNEEDRYEYDR-----YIELNTLKDSLTSN 328

## RESULT 37

US-10-475-204-34  
; Sequence 34, Application US/10475204  
; Publication No. US20050277116A1  
; GENERAL INFORMATION:  
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: PROTEIN INTERACTIONS IN VERTEBRATE CELLS  
; FILE REFERENCE: HMV-056.25  
; CURRENT APPLICATION NUMBER: US/10/475,204  
; CURRENT FILING DATE: 2003-10-17  
; PRIOR APPLICATION NUMBER: PCT/US02/13008  
; PRIOR FILING DATE: 2002-09-27  
; PRIOR APPLICATION NUMBER: 60/285,509  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-475-204-34

Query Match 9.7%; Score 82; DB 6; Length 943;  
Best Local Similarity 23.2%; Pred. No. 49;  
Matches 44; Conservative 18; Mismatches 76; Indels 52; Gaps 7;



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 82.1834 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TTVKEFILNKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	4	US-10-067-385-8
2	848	100.0	2119	3	US-09-769-744A-28
3	848	100.0	2140	4	US-10-282-122A-73670
4	848	100.0	2140	5	US-10-472-928-1180
5	845	99.6	637	5	US-10-617-320-3169
6	615	72.5	117	3	US-09-765-272-68
7	615	72.5	117	6	US-11-106-649-68
8	119	14.0	778	4	US-10-724-972A-5663
9	111.5	13.1	775	4	US-10-282-122A-70721
10	111	13.1	707	4	US-10-282-122A-52942
11	110.5	13.0	647	5	US-10-691-672A-3
12	108	12.7	188	5	US-10-691-672A-7
13	107.5	12.7	470	5	US-10-739-930-6262
14	106	12.5	665	3	US-09-820-843A-107
15	104	12.3	540	5	US-10-732-923-22820
16	103.5	12.2	169	5	US-10-691-672A-2
17	103	12.1	903	4	US-10-282-122A-52328
18	101.5	12.0	564	6	US-11-097-143-12723
19	101.5	12.0	948	5	US-10-732-923-4286
20	101	11.9	973	5	US-10-732-923-18783
21	100.5	11.9	1373	5	US-10-732-923-16976
22	100	11.8	3127	5	US-10-732-923-22588
23	99.5	11.7	1384	4	US-10-473-576-22
24	99.5	11.7	1404	4	US-10-473-576-2
25	97.5	11.5	1184	4	US-10-282-122A-53254
26	97.5	11.5	2468	4	US-10-755-889-615
27	97.5	11.5	2468	5	US-10-489-740-216

28	97.5	11.5	2519	5	US-10-450-763-46995	Sequence 46995, A
29	96.5	11.4	898	4	US-10-425-115-205148	Sequence 205148, A
30	96.5	11.4	2060	4	US-10-381-596A-2	Sequence 2, Appl1
31	96	11.3	1419	5	US-10-732-923-4285	Sequence 4285, Ap
32	95.5	11.3	1350	5	US-10-732-923-8760	Sequence 8760, Ap
33	95.5	11.3	1434	5	US-10-732-923-4235	Sequence 4235, Ap
34	94.5	11.1	3124	5	US-10-732-923-22709	Sequence 22709, A
35	93.5	11.0	645	4	US-10-282-122A-70294	Sequence 70294, A
36	93.5	11.0	645	5	US-10-470-048B-414	Sequence 414, Appl
37	93.5	11.0	654	4	US-10-172-502-10	Sequence 10, Appl
38	93.5	11.0	654	6	US-11-020-509-10	Sequence 10, Appl
39	93.5	11.0	932	4	US-10-282-122A-52510	Sequence 52510, A
40	93	11.0	869	4	US-10-437-963-12282	Sequence 12282, A
41	93	11.0	933	4	US-10-437-963-166606	Sequence 166606, A
42	93	11.0	1529	5	US-10-732-923-8762	Sequence 8762, Ap
43	92.5	10.9	645	5	US-10-470-048B-142	Sequence 142, Appl
44	92.5	10.9	648	4	US-10-425-115-252919	Sequence 252919, A
45	92	10.8	295	4	US-10-282-122A-70503	Sequence 70503, A

## ALIGNMENTS

## RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score	848	DB	4	Length	773
Best Local Similarity	100.0%	Pred. No.	7.3e-64	Mismatches	0	Indels	0
Matches	164	Conservative	0	Gaps	0		
QY	1	TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG	60				
Db	610	TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG	669				
QY	61	WEISGFEKGKAGYVINLSKDTFKVPVKIEEKKEENKPTFDVSKKNDPNQVNHSQLN	120				
Db	670	WEISGFEKGKAGYVINLSKDTFKVPVKIEEKKEENKPTFDVSKKNDPNQVNHSQLN	729				
QY	121	ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK	164				
Db	730	ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK	773				

## RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins



```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 848; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.5e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1922 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 1981
|
|
|
QY 61 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 120
|
|
|
Db 1982 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 2041
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2042 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2085
|
|
|

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 848; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1943 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002
|
|
|
QY 61 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 120
|
|
|
Db 2003 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 2062
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2063 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2106
|
|
|

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472923
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.8+01)
US-10-472-928-1180

Query Match      100.0%; Score 848; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1943 TTVKFPIILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002
|
|
|
QY 61 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 120
|
|
|
Db 2003 WEISGPEGKKGAGYVILNSKDTFIPKVPFKIEEKKKEENKPTFDVSKKKDNPQVNHSQLN 2062
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2063 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2106
|
|
|

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

Publication No. US20050136404A1  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
FOR DIAGNO  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1...637  
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
US-10-617-320-3169  
Query Match 99.6%; Score 845; DB 5; Length 637;  
Best Local Similarity 99.4%; Pred. No. 1e-63; Mismatches 0; Indels 0; Gaps 0;  
Matches 163; Conservative 1;  
QY 1 TTVKFELMKTGVSSELKPHRTVTIQTGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 440 TTVKFELMKTGVSSELKPHRTVTIQTGKEMSTIVSEEDFILPVYKGELEKGYQFDG 499  
QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIEKKEENKPTFDVSKKKDNPQVNHSQLN 120  
DB 500 WEISGFEKGDAGYVNLKDTFIKPVFKIEKKEENKPTFDVSKKKDNPQVNHSQLN 559  
QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164  
DB 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 603  
RESULT 6  
US-09-765-272-68  
Sequence 68, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: Hp Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-272-68  
Query Match 72.5%; Score 615; DB 3; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.1e-45; Mismatches 0; Indels 0; Gaps 0;  
Matches 117; Conservative 0;  
QY 48 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFIKPVFKIEKKEENKPTFDVSK 107  
DB 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFIKPVFKIEKKEENKPTFDVSK 60  
QY 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164  
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117  
RESULT 7  
US-11-106-649-68  
Sequence 68, Application US/11106649  
Publication No. US20050181439A1  
GENERAL INFORMATION:  
APPLICANT: Choi et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
FILE REFERENCE: PB340P2C3D1  
CURRENT APPLICATION NUMBER: US/11/106,649  
CURRENT FILING DATE: 2005-04-15  
PRIOR APPLICATION NUMBER: US 09/765,271  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: US 09/536,784  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: US 08/961,083  
PRIOR FILING DATE: 1997-10-30  
PRIOR APPLICATION NUMBER: US 60/029,960  
PRIOR FILING DATE: 1996-10-31  
NUMBER OF SEQ ID NOS: 454  
SOFTWARE: Patent in version 3.3  
SEQ ID NO 68

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      72.5%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 6.1e-45;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YGELEKGYQPGWEISGEGKDDAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 107
DB 1 YGELEKGYQPGWEISGEGKDDAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 60

QY 108 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNNPK 164
DB 61 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNNPK 117

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US2004014773441
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S.epidermidis
US-10-724-972A-5663

Query Match      14.0%; Score 119; DB 4; Length 778;
Best Local Similarity 27.0%; Pred. No. 0.17;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK-----GELEKGYQDGV-- 61
DB 617 EDVNAQSLKP-----ITGNGKQIKQOSVSGTKVLPKSHKVMMLTDTGELTMP-DMTGWTK 671

QY 62 -BISGFE-----GKKDAGYVIN--LSKDTFIKPVFK-----K 107
DB 672 EDVLAPEDLTKIKVSTKNGFPVTNQISKGQIIL-----NKKIEVLSAED 718

QY 108 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNN 161
DB 719 TDDQEKTDSDSDNKSDDKADSHSNTSSSTKN-----DKSNADSKNDD 766

RESULT 9
US-10-282-122A-70721
; Sequence 70721, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70721
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70721

Query Match      13.1%; Score 111.5; DB 4; Length 775;
Best Local Similarity 24.9%; Pred. No. 0.74;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK-----GELEKGYQDGV-- 61
DB 614 EDVNAQSLKP-----ITGNGKQIKQOSVSGTKVLPKSHKVMMLTDTGELTMP-DMTGWTK 668

QY 62 -EISGFE-----GKKDAGYVIN--LSKDTFIKPVFK-----KIEEKKKEENKPTF 103
DB 669 EDVLAPEDLTKIKVSTKNGFPVTNQISKGQIILKKDKIEVLSAEDTDDQEKTDSDSS 728

QY 104 DVSKKQNPQVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNN 152
DB 729 DKSKCKQKVDNDSNASSSSKNEKSNADSKNDDSDSTNETSGS--ERNN 775

RESULT 10
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

;; FILE REFERENCE: ELITRA.034A

;; CURRENT APPLICATION NUMBER: US/10/282,122A

;; CURRENT FILING DATE: 2003-02-20

;; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: 60/230,335

;; PRIOR FILING DATE: 2000-09-06

;; PRIOR APPLICATION NUMBER: 60/230,347

;; PRIOR FILING DATE: 2000-09-09

;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23

;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/267,636

;; PRIOR FILING DATE: 2001-02-09

;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16

;; Remaining Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 78614

;; SOFTWARE: PatentIn version 3.1

;; SEQ ID NO 52942

;; LENGTH: 707

;; TYPE: PRT

;; ORGANISM: Clostridium difficile

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (6)..(6)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (18)..(18)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (29)..(29)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (37)..(37)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (43)..(43)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (54)..(54)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (84)..(84)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (86)..(86)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (359)..(359)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (385)..(385)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; FEATURE:

;; LOCATION: (388)..(388)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (396)..(396)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (400)..(400)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (402)..(402)

;; OTHER INFORMATION: X=any amino acid

;; FEATURE:

;; NAME/KEY: MISC\_FEATURE

;; LOCATION: (404)..(404)

;; OTHER INFORMATION: X=any amino acid

;; US-10-282-122A-52942

Query Match

Best Local Similarity 13.1%; Score 111; DB 4; Length 707;

Matches 44; Conservative 26; Mismatches 58; Indels 38; Gaps 6;

QY 3 VKRPT--LNKDTGEVSELKPHRVTVTIQNGKMSSTIVSBEEDFILPVYKGELEKGYQPDG 60

Db 494 VKREIDSNQDIDGVDEKQ-----TTDKEDYS---NKEDIIPEPNKSKKKAKLFG- 542

QY 61 WEISGFEGKDGAGYVYNLSKDTFKVPFKLEKKEEN--KPTFDVSKKDNPOVNHQ 118

Db 543 -----FIKONERVEOEENLNDISPDILDKPVENNQVKSEE 580

QY 119 LNESHRKEDLQREHSQKSDTKDVTATVLDKNMISSKSTTNPNK 164

Db 581 IEQNELKE-IKQEPSQHIEERSVKIEKPINNLDKVSNNESK 625

RESULT 11

US-10-691-672A-3

;; Sequence 3, Application US/10691672A

;; Publication No. US20050112133A1

;; GENERAL INFORMATION:

;; APPLICANT: DRUILHE, PIERRE

;; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
;; MALARIAL VACCINES CONTAINING IT

;; FILE REFERENCE: 02356.0085

;; CURRENT APPLICATION NUMBER: US/10/691,672A

;; NUMBER OF SEQ ID NOS: 13

;; SOFTWARE: PatentIn Ver. 3.3

;; SEQ ID NO 3

;; LENGTH: 647

;; TYPE: PRT

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

;; OTHER INFORMATION: Peptide

;; FEATURE:

;; NAME/KEY: SITE

;; LOCATION: (1)..(647)

;; OTHER INFORMATION: GLURP MSP3 fusion protein

;; US-10-691-672A-3

Query Match

Best Local Similarity 13.0%; Score 110.5; DB 5; Length 647;

Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;

QY 4 KEFILNKDTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53

Db 418 EEAVSEKNAHETVB---HEETVQSBNPEKADNGVSNLNELNENFV-----ESE 468

QY 54 K-----GYQPD-GWEISGP--EGKKDAG-----YVNLSDTFIKPVFKIE 92

Db 469 KSEHEARSKAEASSVDYILGWFEGGVPEHKKEENMLSHLYVSSKKENISKENDVDLD 528  
Qy 93 EKKEENKPTFDVSKKONPQVNHSLN-----ESHKEDLQREHHSQKSDS 139  
Db 529 E-KKEAEETEELFEKNEETESISISDEEBEKEEKEKQKEQENNDQ 587  
Qy 140 TKDVTATVLDKNNISSKSTTN 161  
Db 588 KKDEA-----QNLISKQNN 604

RESULT 12  
US-10-691-672A-7  
; Sequence 7, Application US/10691672A  
; Publication No. US20050112133A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIRRE  
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
; FILE REFERENCE: 02356.0085  
; CURRENT APPLICATION NUMBER: US/10/691,672A  
; CURRENT FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patent In Ver. 3.3  
; SEQ ID NO 7  
; LENGTH: 188  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; NAME/KEY: SITE  
; LOCATION: (1)..(188)  
; OTHER INFORMATION: MSP3a to MSP3f  
US-10-691-672A-7

Query Match 12.7%; Score 108; DB 5; Length 188;  
Best Local Similarity 23.2%; Pred. No. 0.26;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
Qy 27 IONGKEMSTIVSEEDFILPVYKGELEKGYQDFGWEISGF--EGKKGAD-----YVINLS 79  
Db 15 VLKAKEASS-----YDIIL-----GWFFGGVPEHKKEENMLSHLYVSSKD 55  
Qy 80 KDTFIKPVFKKIEEKKEE-----ENKPTFDVSKKONPQVNHSLNESHRKE 126  
Db 56 KENISKENDVDLKEEAEATEEELKEEKEEETESEISEDEEBEKEEENKDKK 115  
Qy 127 DLQREHHSQKSDTKDVTATVLDKNNISSKSTTN 161  
Db 116 EQEKEQSNENNDQKDEA-----QNLISKQNN 145

RESULT 13  
US-10-739-930-6262  
; Sequence 6262, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 6262  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270\_1.p  
US-10-739-930-6262

Query Match 12.7%; Score 107.5; DB 5; Length 470;  
Best Local Similarity 20.1%; Pred. No. 0.88;

Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
Qy 19 KPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWEISGFE-----GKK 70  
Db 82 RENRVDTVQNNNGESK-----YQDLARRIRYDE-EATGSAQRIDHPNQK 129  
Qy 71 DAGVINLSKDTFIKPVFKKIEEKKEEENKPTFDVSKKDN----- 111  
Db 130 NVGITEKAPENSPLEETSHRVDDNKRINNQNFTAAKSENNAVSRVSGADHKBAEYMGK 189  
Qy 112 PQVNHSQLNE-----SHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNPNK 164  
Db 190 PMENRDQVRQTESAEKSHRKENVTK3EKPQDEGVKKTAKDKRNEKKEEKTESINK 248

RESULT 14  
US-09-820-843A-107  
; Sequence 107, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 107  
; LENGTH: 665  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|3845248  
US-09-820-843A-107

Query Match 12.5%; Score 106; DB 3; Length 665;  
Best Local Similarity 24.3%; Pred. No. 1.8;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;  
Qy 10 KDTGEVSELKPHRYT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWEI--SGF 66  
Db 127 EEKNKINSLHQNELNLSGK-----NEQDI-----NKKKEKQ-----DISNSNA 169  
Qy 67 EGKKGAGVINLSKDTFIKPVFKKIEEKKEE-----EENKPTFD-----VSKKKNP 112  
Db 170 ENKGD-----VKEGVKELEKKEEKSDDHKVEENKKSDDHKVEENKKSDDH 217  
Qy 113 QVNHSQLNESHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNPNK 164  
Db 218 KVEENKKSDDHKIEVKKVEEHEDEBE-----DKKEKSENKKNKDNK 261

RESULT 15  
US-10-732-923-22820  
; Sequence 22820, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 22820  
; LENGTH: 540  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana



```
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match          12.0%; Score 101.5; DB 6; Length 564;
Best Local Similarity 24.5%; Pred. No. 3.6;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 41 EDPLPVYKGELEKGYQFDG-----EISGPEGKKDAGYVI-----NLSKDTFIK 85
DB 78 EDLDTPLSESRFSK--VPDGVDEHREHDGHVDQPSGEALDDHDEHDDHDDHDEDEE 135

QY 86 PVFKIEEKEENKPT-----FVSKKKDNPQVNHSQLNESHKREDLQREHSQKSDS 139
DB 136 PLTEELEEELEEBEETDEDEPAADVEYEDEDEENNA--GENITAEDAESEEDND 193

QY 140 TKDVTATVLDKNNISKST 158
DB 194 EGTVEATVEATTEAT 212

RESULT 19
US-10-732-923-4286
; Sequence 4286, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4286
; LENGTH: 948
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-4286

Query Match          12.0%; Score 101.5; DB 5; Length 948;
Best Local Similarity 21.8%; Pred. No. 6.8;
Matches 51; Conservative 31; Mismatches 63; Indels 89; Gaps 11;

QY 6 FILNKDTGVSEL-----KPHRVTVTIQNGKEMSTIVSEEDFILPVY-----KG 50
DB 559 YIINLGGVPEELVEDYNAKKHICTLSLOEIKDIOKFLNEBTFKTRGYYSFEKI 618

QY 51 ELEKGYQFDGWEISGPEGKD-----AGVINLSKDTFIKPVFKIEEKEENKPT 102
DB 619 SLAINNSIDHY----FSHMKDNLRVICEPGRYVVAASSTLAVKICKR-----RPT 665

QY 103 F-----DVSKKKDNPQVNHSQLNESHKRED-----LQREHSQK----- 136
DB 666 FQIGIMLKALKAHYDPLNFAQENKKQDEPKINHNNDNNNDNNNNNNNNNNKQGGQ 725
```

```
QY 137 -----SDST-----KQVATATVLDK--NNIS-SKSTNNPN 163
DB 726 GNIMNDLIITSTNDSTNKKNDHSSSQVIQNSCTIRKDEGDNIKINTHTNNPN 779

RESULT 20
US-10-732-923-18783
; Sequence 18783, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18783
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(973)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-18783

Query Match          11.9%; Score 101; DB 5; Length 973;
Best Local Similarity 21.0%; Pred. No. 7.8;
Matches 34; Conservative 43; Mismatches 73; Indels 12; Gaps 4;

QY 7 ILNKDTGE-----VSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
DB 581 LVNNNTNQNGNNSYDDKLETYNINMKDQKGECSYTKT----LIQHRSGSKSGNKFPIH 636

QY 61 WEISGPEGKKDAGVINLSKDTFIKPVFKIEEKEENKPTFVSKKKDNPQVNHSQLN 120
DB 637 TKINNINREHARGYYSLSKDNNEIENVNDNIEVNDNIEVNDNIEVNDNIEVRNDSIN 696

QY 121 ESHRKEDL-QREHSQKSDSTKQVATATVLDKNNISKSTNN 161
DB 697 DONKEKNISETNSPNSKSEYT-FVTATSNKDDNINKSSND 737

RESULT 21
US-10-732-923-16976
; Sequence 16976, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 16976
; LENGTH: 1373
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-16976

Query Match          11.9%; Score 100.5; DB 5; Length 1373;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 6 FILNKDTGVSE----LKPHEVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 61
DB 418 YVNDEDRGVIVSEESTIMLPFHVQIILL-----LSATV-----PNY-----LEFADW 457

QY 62 -----EISGPEGKK-----DAGVINLSKDTFIKPVFK-----KIEEKGEE 98
```



Db 458 VQFTQKQVISTKRPVPLHLYIVYDSVYVMDKRNKFFYSAPKEIYVIRKQBKAN 517  
Qy 99 NKPTFDVSKKONPOVNHSQLNESH--RKEDLQREHSHSQSDSTKDVATVLDKNNISSKS 157  
Db 518 NNTKQITSGSNNTSSNLKKNYYDSKKYLLTNNKENDNT-----QNNNNNN 568  
Qy 158 TTNV 161  
Db 569 NNNN 572

## RESULT 22

US-10-732-923-22588  
; Sequence 22588, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; PRIOR FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 22588  
; LENGTH: 3127  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum 3D7  
US-10-732-923-22588

Query Match 11.8%; Score 100; DB 5; Length 3127;  
Best Local Similarity 24.2%; Pred. No. 40;  
Matches 52; Conservative 34; Mismatches 71; Indels 58; Gaps 9;  
Qy 4 KEPILKDTGVSSELKPHRVTVTIQNG-----KEMSSSTIVSEEDFILPVYKGELEKGVQ 57  
Db 1969 KKF--KRNTSYVLESPLHLIGDIVDNNIKRKKKKKEIKTIVSDDMFTSPVNIKEYNYNEQ 2026  
Qy 58 PDGWEISG--PEGKKDAGYVNLKDTPIKPVFKKIEKKEEENK-----100  
Db 2027 ERKKEIVGNLSYDKTKIIPFKFTKEGRKK--KKIEKKEKKEKENNNFLYNDYSS 2084  
Qy 101 ---PTF-----DVSKKONPOVNHSQL-----NESHK---EDLQR 130  
Db 2085 YSSPKYGDNNNFVIKIRERKDFQKCPDHPNPNFSEKFLHYNPMKNKKNKNNKNNVR 2144  
Qy 131 EHSQKSDTKD--VTATVLDKNNISSKSTTNNPK 164  
Db 2145 NEYPNTSSKDGVSYNFLSDSLFSSDNEYSSDNE 2179

## RESULT 23

US-10-473-576-22  
; Sequence 22, Application US/10473576  
; Publication No. US20040101884A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION  
; APPLICANT: LU, DYUNG AINA M.  
; APPLICANT: ARVIZU, CHANDRA S.  
; APPLICANT: GANDHI, AMEENA R.  
; APPLICANT: HAFALIA, APRIL J.A.  
; APPLICANT: DING, LI  
; APPLICANT: LU, YAN  
; APPLICANT: RAMKUMAR, JAYALAXMI  
; APPLICANT: SWARNAKAR, ANITA  
; APPLICANT: TANG, Y. TOM  
; APPLICANT: YUE, HENRY  
; APPLICANT: TRAN, BAO  
; APPLICANT: LEE, SOO YUEN  
; APPLICANT: WARREN, BRIDGET A.  
; APPLICANT: NGUYEN, DANNIEL B.  
; APPLICANT: THANGAVELU, KAVITHA

; APPLICANT: YAO, MONIQUE G.  
; APPLICANT: ELLIOTT, VICKI S.  
; APPLICANT: BAUGHN, MARIAH R.  
; APPLICANT: EMERLING, BROOKE M.  
; APPLICANT: LAL, PREETI G.  
; APPLICANT: GIETZEN, KIMBERLY J.  
; APPLICANT: BECHA, SHANYA D.  
; APPLICANT: MARQUIS, JOSEPH P.  
; APPLICANT: KABLE, AMY E.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
; FILE REFERENCE: PP-0921 USN  
; CURRENT APPLICATION NUMBER: US/10/473,576  
; CURRENT FILING DATE: 2003-09-29  
; PRIOR APPLICATION NUMBER: PCT/US02/09809  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: US 60/280,387  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: US 60/282,335  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/286,663  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: US 60/285,484  
; PRIOR FILING DATE: 2001-04-19  
; PRIOR APPLICATION NUMBER: US 60/350,702  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/351,749  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PERL Program  
; SEQ ID NO 22  
; LENGTH: 1384  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 7506096CD1  
US-10-473-576-22

Query Match 11.7%; Score 99.5; DB 4; Length 1384;  
Best Local Similarity 21.7%; Pred. No. 16;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
Qy 4 KEPILKDTGVSSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGVQF-----58  
Db 1013 KEPTMLQNEQISQLK-KBIERTQORMKEMSVNKBQEQIATQYKRAIDLQELRLTRE 1071  
Qy 59 -----DGWEISGFEKKDA-----GYVINLSK-----80  
Db 1072 QVNSHTELAEARHQVQAQREIERLSSELEDMKQLSKEDAHGNHNLABELGASKVREAH 1131  
Qy 81 -----DTPIKPVFKKIEKKE-----EENKPTDVSKKONPOVNHSQLNESHK 126  
Db 1132 LEARMQAIEKKLSAEVESLKEAYHMEMISHQENHAKWKIS--ADSKSVQQLNEQLEKA 1189  
Qy 127 DLQREHSHSQSDSTKDVATVLDKNNI 153  
Db 1190 KLELEB---AQDTVSNLHQVQDRNEV 1213

## RESULT 24

US-10-473-576-2  
; Sequence 2, Application US/10473576  
; Publication No. US20040101884A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION  
; APPLICANT: LU, DYUNG AINA M.  
; APPLICANT: ARVIZU, CHANDRA S.  
; APPLICANT: GANDHI, AMEENA R.  
; APPLICANT: HAFALIA, APRIL J.A.  
; APPLICANT: DING, LI  
; APPLICANT: LU, YAN  
; APPLICANT: RAMKUMAR, JAYALAXMI  
; APPLICANT: SWARNAKAR, ANITA

APPLICANT: TANG, Y. TOM  
APPLICANT: YUB, HENRY  
APPLICANT: TRAN, BAO  
APPLICANT: LEE, SOO YUEN  
APPLICANT: WARREN, BRIDGET A.  
APPLICANT: NGUYEN, DANNIEL B.  
APPLICANT: THANGAVELU, KAVITHA  
APPLICANT: YAO, MONIQUE G.  
APPLICANT: ELLIOTT, VICKI S.  
APPLICANT: BAUGHN, MARIAH R.  
APPLICANT: EMERLING, BROOKE M.  
APPLICANT: LAL, PREETI G.  
APPLICANT: GIETZEN, KIMBERLY J.  
APPLICANT: BECHA, SHANYA D.  
APPLICANT: MARQUIS, JOSEPH P.  
APPLICANT: KABLE, AMY E.  
TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
FILE REFERENCE: PF-0921 USN  
CURRENT APPLICATION NUMBER: US/10/473,576  
PRIOR FILING DATE: 2003-09-29  
PRIOR APPLICATION NUMBER: PCT/US02/09809  
PRIOR FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: US 60/280,387  
PRIOR FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: US 60/282,335  
PRIOR FILING DATE: 2001-04-05  
PRIOR APPLICATION NUMBER: US 60/286,663  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: US 60/285,484  
PRIOR FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: US 60/350,702  
PRIOR FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: US 60/351,749  
PRIOR FILING DATE: 2002-01-25  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 1404  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
FEATURE:  
OTHER INFORMATION: Incyte ID No: 3125036CD1  
US-10-473-576-2

Query Match 11.7%; Score 99.5; DB 4; Length 1404;  
Best Local Similarity 21.7%; Pred. No. 16;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
Qy 4 KEPIILKDTGEVSELKPHRVTTIQKEMSSSTIVSEEDFILPVYKGELEKGYQF----- 58  
Db 1033 KEFIMLQNEQISQLK-KEIERTQORMKEMESVMKEQEQYIATQYKEAIDLGQELRLTRE 1091  
Qy 59 -----DGEIISGPEGKKDA-----GVINLSK----- 80  
Db 1092 QVQNSHTLEAARHQVQARIERLSSELEDKQLSKEKDAHGNHABELGASKVREAH 1151  
Qy 81 -----DTFTKPVFKITEKE-----BENKPTFDVSKKKDNPQVNHLSHSHRKE 126  
Db 1152 LEARMQARIKCLSAEVELKAEYHMEMISHQENHAKWKIS--ADSKSGSVQQLNEQLKA 1209  
Qy 127 DLQREHSQKSDTKDVTATVLDKNNI 153  
Db 1210 KLELEE---AQDTVSNLHQVQQRNEV 1233

## RESULT 25

US-10-282-122A-53254  
Sequence 53254, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 53254  
LENGTH: 1184  
TYPE: PRT  
ORGANISM: Clostridium difficile  
US-10-282-122A-53254

Query Match 11.5%; Score 97.5; DB 4; Length 1184;  
Best Local Similarity 26.3%; Pred. No. 20;  
Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
Qy 14 EVSELKPHRVTTIQ-NGKEMSSSTIV--SEEDF--ILPVYKGELEKGYQFDGWEISGPEG 68  
Db 240 ELSEVNEHRKVIKELNEKEQKVVEKQEDINKEVEVLQDVIEKSVDIYN-SIKGVIS 298  
Qy 69 KQDAGVYNLSKDTF-----IKPVFKIIEKKE-----BENKPTFDVSKKKD 110  
Db 299 KKESS--QINLTKERINFTNEISRNLKIDKEKLNENKQYIKELESNK--LSGSEBLS 354  
Qy 111 NPQVNHLSHSHRKEDLQREHSCSDSTKDVATVLD-KNNISSKSTTTNNPK 164  
Db 355 TLQENIKVLEGSKKQKIKLESNLNNEELLEKESIIDLNKKQFENKLTSLNANK 409

RESULT 26  
US-10-755-889-615  
Sequence 615, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14

```

Query Match      11.4%; Score 96.5; DB 4; Length 898;
Best Local Similarity 24.1%; Pred. No. 17;
Matches 42; Conservative 34; Mismatches 73; Indels 25; Gaps 7;

QY 7 ILNKDTGVSEIKPHRVVTTIQNGKMSSTIVSEEDF-----ILPVYGELEK--- 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 557 ILLKMTTEBSVVA--AASTATEKGKEQABDILEEDBFQDILGQGLTDAKKAELKKCAI 614  
Qy 55 --GYQFDGWEISGF-EGKXGAGVYNLSKDTFIKPVFKKIEKKEEENKPTF-DVSKKKD 110  
Db 615 ACQYKPGATLFGVNEGK-----LRSFROSILLSKVLKMQODLEEEKNAVIOQLAEKVE 669  
Qy 111 NPQVNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSKSTTNNPNK 164  
Db 670 NYEASLKKXGFTIQLLEINVKHEGALEKDFVIQTM--EGSLAEVQTEENKLG 721

## RESULT 30

US-10-381-596A-2  
; Sequence 2, Application US/10381596A  
; Publication No. US20040014178A1  
; GENERAL INFORMATION:  
; APPLICANT: Biostapro AB  
; TITLE OF INVENTION: von Willebrand factor-binding proteins from  
; FILE REFERENCE: 110059600  
; CURRENT APPLICATION NUMBER: US/10/381,596A  
; CURRENT FILING DATE: 2003-07-02  
; PRIOR APPLICATION NUMBER: SE 0003573-3  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2060  
; TYPE: PRT  
; ORGANISM: Staphylococcus lugdunensis  
US-10-381-596A-2

Query Match 11.4%; Score 96.5; DB 4; Length 2060;  
Best Local Similarity 25.7%; Pred. No. 48;  
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

Qy 14 EVSELKPHRVTVTIO-NGKEMSTIVSEBDFILPVYKGELEGYQFDG--WEISGFEKK 70  
Db 1914 DDELK---ITIVDTNGRE-----IVPSKGGQLPP-EQFIGQDWQYGHK--- 1955  
Qy 71 DAGVYNLSKDTFIKPVFKPIEE-----KKEENKPTFDVSKKKNPQVNHSQLNESH 123  
Db 1956 -----IEKDGITTVYKKVENAVPAKQKTKEN--TQSESQPKHTPVQVKQLVKYHN 2006  
Qy 124 RKE--DLQREHSHQSDSTKDVATVLDKNNI 153  
Db 2007 VKEQRSIEKSEHTDMHVSLELPETGETANKGL 2038

## RESULT 31

US-10-732-923-4285  
; Sequence 4285, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 4285  
; LENGTH: 1419  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-732-923-4285

Query Match 11.3%; Score 96; DB 5; Length 1419;  
Best Local Similarity 21.9%; Pred. No. 33;  
Matches 49; Conservative 36; Mismatches 69; Indels 70; Gaps 11;

Qy 6 FILNKDTGVSSEL-----KPHRVTVTIQNGKEMSTIVSEBDFILPVY-----KG 50

Db 1031 YIINLGSGYPEELEVONAKKHKDHIYCTLSIQEIKDUIQKFLNEETPLTKYGYSPFKI 1090  
Qy 51 BLEKGYQFDGWEISGFEKKD-----AGVYNLSKDTFIKPVFKK-----IEEK 94  
Db 1091 SLAINMSTIDHY-----FSHMKDNLRVCEPGRYVMAAGSTLAVKIIGKRPRPTFOGIMLKEL 1146  
Qy 95 KEEENKPTF--DVSKKKDNPOVNH---SQLNESHKEDLQREHSHQ----- 136  
Db 1147 KQHYDPLNFAQENKQKQDETAKINENNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDLIIT 1206  
Qy 137 --SDST-----KQVATVLDK--NNIS-SKSTTNNPN 163  
Db 1207 STNDSTKKNHSSSQVQNVSCITIRKDEGDNINIKINTINNPN 1250

## RESULT 32

US-10-732-923-8760  
; Sequence 8760, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 8760  
; LENGTH: 1350  
; TYPE: PRT  
; ORGANISM: Plasmodium yoelii yoelii  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1350)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-732-923-8760

Query Match 11.3%; Score 95.5; DB 5; Length 1350;  
Best Local Similarity 23.2%; Pred. No. 34;  
Matches 45; Conservative 35; Mismatches 71; Indels 43; Gaps 10;

Qy 1 TTVKEFILNKDTGVSSELKPHRVTVTIQNGK-----EMSSTIVSEBDFILPVYKGELEK 54  
Db 852 STIRKFLKNV-----KGMIFLDLGERSKKVMWNTACTTKTKKAILYGHKAEARG 903  
Qy 55 GYQFDG-----EISGFEKKDAGVYNLSKDTFIKPVFKKIEKKEE---NKPTPDV 105  
Db 904 GHDFQFNNKLDLEKIEEKWN-NYHINQK---IKEIISQTEKEDFEKIIINHEFLT 959  
Qy 106 SKKKD-----NPQV-NHSQLNESHKEDLQRE-----EHSQKSDSTKDVATVLDK 150  
Db 960 HHAKNIVNLINLPNINLDSNIGDSIEEKDVSKDIOEDGNEGKYGKSGFNGI-VNIYSK 1018  
Qy 151 NNISKSTTNNPNK 164  
Db 1019 SNVGSSETRNDSK 1032

## RESULT 33

US-10-732-923-4235  
; Sequence 4235, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149

```
; SEQ ID NO 4235
; LENGTH: 1434
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-4235

Query Match
Best Local Similarity 11.3%; Score 95.5; DB 5; Length 1434;
Matches 52; Conservative 31; Mismatches 70; Indels 75; Gaps 12;

QY 6 FILNKDTGEVSEL-----KPHRVTVTQNGKEMSTIVSEEDFILPVY-----KG 50
Db 1043 YIINLGGVPELEVDNAKHKDIHYCTLSLQIKDKQKPLNEETFLTKKYGYSPKI 1102
QY 51 ELEKGYQDFGWISGFGKKO-----AGVINLSKDTFTFK-----PVFKKI----- 91
Db 1103 SLAINKSIDHY-----FSHKMLRVICBPGRWAAASSTLAVKIIIGKRPTFGIMLKDL 1158
QY 92 -----BEKKEBENKTFP--DVSKKDNQVNHSQLNESHKKE-----DL 128
Db 1159 KDHYDPLNPAQENKKQDETIKHNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDL 1218
QY 129 -----OREBHSQKSDTSKQVATVLDK--NNIS-SKSTTNPN 163
Db 1219 IITSTNDSTNKNDHS--SSQVIONVSCIRDEGDKINHTINPN 1265

RESULT 34
US-10-732-923-22709
; Sequence 22709, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22709
; LENGTH: 3124
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-22709

Query Match
Best Local Similarity 11.1%; Score 94.5; DB 5; Length 3124;
Matches 50; Conservative 33; Mismatches 74; Indels 55; Gaps 8;

QY 4 KEFILNKDTGEVSELKPHRVTVTQNG-----KEMSTIVSEEDFILPVYKGELEKGYQ 57
Db 1969 KKF--KRNTSVYLESPLHLIGDIVDNNIKRKKKKKEIKTIVSDDMFTSPVNIKEVYNEQ 2026
QY 58 FDGWISG---PEGKQAGVINLSKDTFTFKPVFKKIEKKEENKPTF-----103
Db 2027 ERKKEIVGNLSYDKTKKICPPIKFTKEGRIKK--NKIEKKEKKEYNFNFLYNDYSSYSS 2084
QY 104 -----DVSKKDNQVNHSQL-----NESHKKE---DLQREH 133
Db 2085 PKYGNENNFTVITYREKDFQKFDHPNFNFKPLHNYNPMKNQKNQKNVVRNEY 2144
QY 134 SQKSDSTKD-VTATVLDKNNISKSTTNPNK 164
Db 2145 PNYTSSKQGVSYNLFSLSDLSFSSDNEYSSDNE 2176

RESULT 35
US-10-282-122A-70294
; Sequence 70294, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
```

```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvakind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70294
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-70294

Query Match
Best Local Similarity 11.0%; Score 93.5; DB 4; Length 645;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFILNKDTGR-VSELKPHRVTV--TQNGKEMSTIVSEEDFILPVYKGBLE-KGY 56
Db 327 SAITFQNVQPTNEKMTDLQDTKYVYVESVENNESMDTFVKH-----PIKTGMLNGKKY 381
QY 57 QP-----DGWEISGFGKK-----D 71
Db 382 MVMETTNDNDYKQPMVEGQVRVTISKDAKNKTRTIIPYVEGKTYLDAIVKRVHVKTIIDYD 441
QY 72 AGVYINL-SKDTFTKPVFKKIEKKEENKPTFDV-----SKKDNQVNHSQLNESHK 125
Db 442 GQYHVRIVDEAFKANTDKNSKKEQQDNNAKKEATATPSKPTSPVEKESQKDSQK 501
QY 126 EDLQ-----REEHSQKSDTKDVT-ATVLDKNNISKSTTNPNK 164
Db 502 DNKQLPSVEKENDASSESGKDKTPTATKPTKGEVSSSTI--PTK 543

RESULT 36
US-10-470-048B-414
; Sequence 414, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; ANTIGENS TO A SPECIFIC PATHOGEN
```

```

; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470.048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 414
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-414

Query Match 11.0%; Score 93.5; DB 5; Length 645;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11.

Qy 1 TTVKEFILNKDTGE-VSELKPHRVTV--TIQNGKEMSTIYSEEDFILPVYKGELE-KGY 56
Db 327 SAITEFQNVQPTNKMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKY 381
Qy 57 QF-----DGWEISGFEGKK-----D 71
Db 382 MVMETTTNDYDQDFWVEGQVRVTISKAKNNTRTIIFPVYEGKTIYDAIVKVHVKTIIYD 441
Qy 72 AGYVINL-SKDTFIKPVFKLEEKKEEENKPTFDV-----SKKKDNQPWNHSQLNESHRK 125
Db 442 GQYHVRIVDKEAFTKANTDKSNKKEQQDSAKCEATPATPSKPTSPVYKESQKQDSQKD 501
Qy 126 EDLQ-----REBHSQKGSOSTKQV--ATVLDKNNISKSTTNNPNK 164
Db 502 DNKQLPSVKEKNDASSGKDKTPATKPTKGEVSSSTT--PTK 543

```

```

RESULT 37
US-10-172-502-10
; Sequence 10, Application US/10172502
; Publication No. US20030185833A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES.
; FILE REFERENCE: P072630S01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

```

Query Match	11.0%;	Score 93.5;	DB 4;	Length 654;
Best Local Similarity	21.9%;	Pred. NO. 21;		
Matches	49;	Conservative 34;	Mismatches 74;	Indels 67; Gaps 111
Qy	1	TTVKEFILNKDTGB-VSELKPHRVTV--TIOGKMSSTIVSEEDFILPVYKGELE-KGY	56	
Db	336	SAITEFQNVQPTNEKWTDLQDTKYVYVESVENNESMDTFFVKH-----PIKTGMLNGKKY	390	
Qy	57	QF-----DGWEISGFEGKK-----	71	
Db	391	MMVETTTNDYDKDFMVEGQQRVRTISKAKNNTRTIIPPVYEGTKLYDAIVKVHVKTIDYD	450	
Qy	72	AGVYINL-SKDTFTKPVPKLEEKEEENKPTFDV-----SKKDNQPNVHSQLNESHK	125	
Db	451	GOYHVRIVDVKAFYKANTDKNKKEQQDNASKAEATPATPSKPTSPVEKESQKQDSQKD	510	
Qy	126	EDLQ-----REESHQSQSDTKDVT-ATVLDKNNISKSSTNNPNK	164	
Db	511	DNKOLPSVEKENDASSGKKDTKATKPTKGEVESSTT--PTK	552	

```

RESULT 38
US-11-020-509-10
; Sequence 10, Application US/11020509
; Publication No. US20050106648A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US02/BAS
; CURRENT APPLICATION NUMBER: US/11/020,509
; CURRENT FILING DATE: 2004-12-27
; PRIOR APPLICATION NUMBER: US 10/172,502
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-11-020-509-10

Query Match 11.0%; Score 93.5; DB 6; Length 654;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

Qy 1 TTVKBFILNKDTGE-VSELKPHRVTV--TIOGKEMSTIVSEEDFILPVYKGLR-KGY 56
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 SAITEFQNVQPTNEKMTDLQDTKYV/VYESVENNESMMDTFVKH-----PIKTGMLNGKKY 390

Qy 57 QF-----DQWELSGEGKK-----D 71
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 MVMETTNDDYDKDFMVEGQVRVTSIKAKNTRTIIPVYEGKLYDAIVKVVHTIDYD 450

Qy 72 AGVYINL-SKDTFIKPVFKIEBKSEENKPTDV-----SKKKNPQVNHSQLNESHK 125
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GQYHVRIVDXEATKANTDKSNKKEQQDSAKKEATPATPSKTPSPVEKESQKDSQKD 510

Qy 126 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISSKSTTNNPNK 164
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 DNKOLPSVKENDASSESGKDTPATPKTKGEVSSSTT--PTK 552
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 39  
US-10-282-122A-52510  
; Sequence 52510, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347

Search completed: April 24, 2006, 15:43:36  
Job time : 83.1834 secs

```
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 52510
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52510

Query Match      11.0%; Score 93.5; DB 4; Length 932;
Best Local Similarity 24.4%; Pred. No. 32;
Matches 42; Conservative 30; Mismatches 53; Indels 47; Gaps 9;

QY      8 LNKDGEVSELKPHRTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFE 67
DB      735 LTKLEDNIEGVNYSVS-----ELENEIV----FLRKIKGGADQSY-----GTE 777

QY      68 GKGDAGY---VINLSKDTIKVPFKIEKKE-----ENKPTFDVSKKKN-- 111
DB      778 VAKLAGLPSPVINRAKE-----ILQHIEGDKKEENSINTAPSKYKDYIEVSKDTSNTK 832

QY      112 ----PQVNHSQLNESHRK---EDLQREHSQKSDSTKVATVLDKNMISSK 156
DB      833 NNLGSEIKDITLSETNTDTIIEDESTKEH--LSSNKKQINCRINDEKSIKKB 882

RESULT 40
US-10-437-963-12282
; Sequence 12282, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25224C.1.pap
US-10-437-963-12282

Query Match      11.0%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 33;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY      104 DVSKKKNPQVNHSQLNESHRKEDLQREHSQKSDSTKVDTA--TVLDKNN 152
DB      617 DASKKDNHQSEGNL--SHRDEDPTRKKKQKTKNATSDACAQEVVTEKNN 665
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 25.4989 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVKEFILNKDTGEVSELKP.....ATVLDKNKISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgm2\_6/ptodata/1/iaa/5 COMB.pap:\*  
2: /cgm2\_6/ptodata/1/iaa/6 COMB.pap:\*  
3: /cgm2\_6/ptodata/1/iaa/H COMB.pap:\*  
4: /cgm2\_6/ptodata/1/iaa/PTCUS COMB.pap:\*  
5: /cgm2\_6/ptodata/1/iaa/RE COMB.pap:\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles.1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	2	US-09-590-991-8
2	845	99.6	637	2	US-09-107-433-3169
3	845	99.6	2138	2	US-09-583-110-5274
4	615	72.5	117	2	US-08-961-083-68
5	615	72.5	117	2	US-09-536-784-68
6	615	72.5	117	2	US-09-765-271-68
7	615	72.5	117	2	US-09-765-272A-68
8	119	14.0	746	2	US-09-710-279-652
9	119	14.0	778	2	US-09-134-001C-3868
10	101.5	12.0	347	2	US-09-248-796A-16224
11	97.5	11.5	2468	2	US-09-576-594-726
12	97.5	11.5	2468	2	US-09-538-092-1135
13	97.5	11.5	2522	2	US-09-949-016-10237
14	95	11.2	348	2	US-09-538-092-1316
15	93.5	11.0	654	2	US-10-172-502-10
16	93	11.0	280	2	US-09-248-796A-17646
17	92	10.8	299	2	US-09-710-279-1888
18	92	10.8	309	2	US-09-134-001C-5667
19	91	10.7	442	2	US-09-134-001C-3033
20	91	10.7	902	2	US-09-134-001C-5157
21	90.5	10.7	743	2	US-08-910-925-3
22	90.5	10.7	743	2	US-09-949-016-6261
23	90.5	10.7	758	2	US-09-949-016-8288
24	89.5	10.6	262	2	US-09-248-796A-21451
25	89.5	10.6	402	2	US-09-464-483-4
26	89.5	10.6	402	2	US-09-414-664-4
27	89.5	10.6	511	2	US-09-198-452A-509

28	89.5	10.6	511	2	US-09-438-185A-475	Sequence 475, Appl
29	89.5	10.6	529	2	US-09-464-483-2	Sequence 2, Appli
30	89.5	10.6	529	2	US-09-414-664-2	Sequence 2, Appli
31	89.5	10.6	553	2	US-09-248-796A-16588	Sequence 16588, A
32	87.5	10.3	825	2	US-09-248-796A-16538	Sequence 16538, A
33	87.5	10.3	1315	2	US-09-200-650E-5	Sequence 5, Appli
34	87.5	10.3	1702	2	US-08-296-731-5	Sequence 5, Appli
35	87.5	10.3	1702	2	US-09-839-996-5	Sequence 5, Appli
36	87.5	10.3	1702	2	US-10-080-505-5	Sequence 5, Appli
37	87.5	10.3	1702	2	US-10-645-655-5	Sequence 5, Appli
38	87.5	10.3	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
39	86.5	10.2	277	2	US-09-830-230A-651	Sequence 651, App
40	86.5	10.2	732	1	US-08-533-669A-18	Sequence 18, Appl
41	86.5	10.2	732	2	US-09-183-861-18	Sequence 18, Appl
42	86.5	10.2	732	2	US-09-022-765-18	Sequence 18, Appl
43	86.5	10.2	732	2	US-09-551-974A-18	Sequence 18, Appl
44	86.5	10.2	732	2	US-09-565-501A-18	Sequence 18, Appl
45	86.5	10.2	732	2	US-09-639-206A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590.991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%	Score 848;	DB 2;	Length 773;
Best Local Similarity	100.0%	Pred. No. 6.2e-79;		
Matches 164;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEDDFILPVYKGBLEKGYQFDG	60	
Db	610	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEDDFILPVYKGBLEKGYQFDG	669	
Qy	61	WEISGFEKGDAGVYINLSKDTFKPVPKLEEKKEBENKPTFDVSKKKNPQVNHSQLN	120	
Db	670	WEISGFEKGDAGVYINLSKDTFKPVPKLEEKKEBENKPTFDVSKKKNPQVNHSQLN	729	
Qy	121	ESHRKEDIQREHSQKSDSTKDVATVLDKKNISSKSTNNPNK	164	
Db	730	ESHRKEDIQREHSQKSDSTKDVATVLDKKNISSKSTNNPNK	773	

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; FOR DIAG

```
/
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD/ROM ISO9660
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/
/ CURRENT APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION NUMBER: 60/ 085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 3169:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 637 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATON 1...637
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
/
/ US-09-107-433-3169
/
/ Query Match 99.6%; Score 845; DB 2; Length 637;
/ Best Local Similarity 99.4%; Pred. No. 9.8e-79;
/ Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 TTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDG 60
/ DB 440 TTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDG 499
/
/ QY 61 WEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 120
/ DB 500 WEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 559
/
/ QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
/ DB 560 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 603
/
/ RESULT 3
/ US-09-583-110-5274
/ Sequence 5274, Application US/09583110
/ Patent No. 6699703
/
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al.
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
/ FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
/ FILE REFERENCE: PATH00-07A
/ CURRENT APPLICATION NUMBER: US/09/583,110
/ CURRENT FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/107,433
/ PRIOR FILING DATE: 1998-06-30
/ PRIOR APPLICATION NUMBER: US 60/085,131
/ PRIOR FILING DATE: 1998-05-12
/ PRIOR APPLICATION NUMBER: US 60/051,553
/
```

```
/
/ PRIOR FILING DATE: 1997-07-02
/ NUMBER OF SEQ ID NOS: 5322
/ SEQ ID NO 5274
/ LENGTH: 2138
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ US-09-583-110-5274
/
/ Query Match 99.6%; Score 845; DB 2; Length 2138;
/ Best Local Similarity 99.4%; Pred. No. 5.2e-78;
/ Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 TTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDG 60
/ DB 1941 TTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDG 2000
/
/ QY 61 WEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 120
/ DB 2001 WEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 2060
/
/ QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
/ DB 2061 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104
/
/ RESULT 4
/ US-08-961-083-68
/ Sequence 68, Application US/08961083
/ Patent No. 6159469
/
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340F2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-08-961-083-68
/
/ Query Match 72.5%; Score 615; DB 2; Length 117;
/ Best Local Similarity 100.0%; Pred. No. 5.6e-56;
/ Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 48 YKGELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSK 107
```

```
Db 1 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 60
Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 72.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.6e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 107
Db 1 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 60

Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452

Db 1 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 60
Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68

Query Match 72.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.6e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 107
Db 1 YGELKGYQPDGWEISGFEKGKQAGYVNLNKSOTFIKPVFKIEKKKEENKPTFDVSK 60

Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```



```
Db 260 PVFKILDQRE-----WKAEREQANPKKEBENLNQKPVAKQKQPNSTKQKQYQKQK 314
QY 140 TKDVTATVLDKNNISSKSTNNPNK 164
Db 315 TKKITPKTSKRMLEGISTSNINK 339

RESULT 11
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

Query Match 11.5%; Score 97.5; DB 2; Length 2468;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 584 EKVMVKDKPKVKTETKPSVTEKPSKEEPS-----PV-KAEVA-----EK 623
QY 64 SGFEGKKGAGYVINLSKDTPIKPVFKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDTKPI---KCEBKP 675

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqformatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
```

```
; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match 11.5%; Score 97.5; DB 2; Length 2468;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 584 EKVMVKDKPKVKTETKPSVTEKPSKEEPS-----PV-KAEVA-----EK 623
QY 64 SGFEGKKGAGYVINLSKDTPIKPVFKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDTKPI---KCEBKP 675

RESULT 13
US-09-949-016-10237
; Sequence 10237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10237
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10237

Query Match 11.5%; Score 97.5; DB 2; Length 2522;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 638 EKVMVKDKPKVKTETKPSVTEKPSKEEPS-----PV-KAEVA-----EK 677
QY 64 SGFEGKKGAGYVINLSKDTPIKPVFKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 678 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDTKPI---KCEBKP 729

RESULT 14
US-09-538-092-1316
; Sequence 1316, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformat Version 0.9
; SEQ ID NO 1316
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number Q14093
US-09-538-092-1316

Query Match      11.2%; Score 95; DB 2; Length 348;
Best Local Similarity 30.9%; Pred. No. 0.13;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 49 KGELEKGYQFDGWEISGFGKQDAGYVNLSTQTIKPVFKIIEKKEENKPTP---DV 105
Db 205 ESEGEKG-----GTEKDSKGGKDS-----KKGKDSAIELQAVKADSKDEGDKKDKANGDE 256
QY 106 SK--KKDNPOVNHSLN-----BSHRKEDLQREHSOKSDTKD---VTATVLDKNNI 153
Db 257 SKDAKKDAKEIKKGKDKKKPSSTDSKDDVKKE---SKDQTKDAKKVAKOTEKESA 313
QY 154 SSK 156
Db 314 DSK 316

RESULT 15
US-10-172-502-10
; Sequence 10, Application US/10172502
; Patent No. 6841154
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.0%; Score 93.5; DB 2; Length 654;
Best Local Similarity 21.9%; Pred. No. 0.46;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFILNKDGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
Db 336 SAITEFQNVQPTNEKMTLDQTKYVYVESVNNESNMDTFVKH-----PIKGTMLNGKKY 390
QY 57 QP-----PGWEISGPEGKK-----D 71
Db 391 MYMETNDDYDKDFMVEGQVRTISKAKNRTTIFPVYEGKTYDAIVKVVHTIDYD 450
QY 72 AGVINL-SKDTFIKPVFKIIEKKEENKPTFDV-----SKKKNPQVNHSLQNESHRK 125
Db 451 GQHVIVLNDKEAFTKANTDKSNKEQODNSAKKEATPATPSKPTSPVKEKSKQDSQKD 510
QY 126 EDLQ-----REHSQKSDSTQVTV--ATVLDKNNISSKSTNNPNK 164
Db 511 DNKQLPSVKEKNDASSSESGDKTTPATKPTKGEVSSSTT--PTK 552
```

## RESULT 16

```
US-09-248-796A-17646
; Sequence 17646, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17646
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17646
```

```
Query Match      11.0%; Score 93; DB 2; Length 280;
Best Local Similarity 25.7%; Pred. No. 0.16;
Matches 43; Conservative 26; Mismatches 58; Indels 40; Gaps 10;

QY 8 LNKDTGEVSELKPHRVTVTIQNGKEM4MSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFE 67
Db 67 INPIVGHVSS-----TYTVKT--SVASTFCSKYDFNVFSYASNLGLG-----ELYSYA 113
QY 68 GKDDAGYVNLSTQTIKPVFKIE-EKKEENKPTFDVSKKKNPQVNHSLQNESHRK 125
Db 114 NK-----KNSF--PSFEHHEIHSSSEENK-----YKQHPQLQRRHNLHNLHQR 158
QY 126 EDLQREHS-----QKSDSTKCVTATVLDKNNISSKSTNNPN 163
Db 159 VPIKSHKYEGNRTIINPIQLNDNVNPTLLSSNG-STSTTTNNEN 204
```

## RESULT 17

```
US-09-710-279-1888
; Sequence 1888, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1888
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1888
```

```
Query Match      10.8%; Score 92; DB 2; Length 299;
Best Local Similarity 24.4%; Pred. No. 0.22;
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;

QY 9 NKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFI---LPVYKGELEKGYQFDGWEI 63
Db 33 NKDT-EKSDKKYHRIISLIPSTELIYKLGIGEDIVGVSTVDDYDPKDVKKGKKQFDANL 91
QY 64 SGFE-----GKKDAGYVNLSTQTIKPV-----FKKIEKK 95
```



```
Db 92 NKEELIKAKPDILILAHESQKNSAGKVLKSLKDKGVKVVVVKDAQSIDETDTFKSIGQLT 151
QY 96 ERENKPTFDVSKKDNQPNVHSQLNESHKEDLQREHSQ 135
Db 152 DREKQAKELVDTEKENVEKIINSVPKHKKQEVFNEVSSK 191

RESULT 18
US-09-134-001C-5667
; Sequence 5667, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5667
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5667

Query Match 10.8%; Score 92; DB 2; Length 309;
Best Local Similarity 24.4%; Pred. No. 0.23; Mismatches 60; Indels 34; Gaps 5;
Matches 39; Conservative 27;

QY 9 NKTGCVSELKPHRVTTIQNGKMSSTIVSEEDFI----LPVYKGELEKG-YQPDGWEI 63
Db 43 NKDT-EKSDKKVRIISLPSNTYILYRLGIGEDIVGVSTVDYDFDKVKGKKQFDANWL 101
QY 64 SGFE-----GKKDAGYVINLSKDTPIKPV-----FKKIBKK 95
Db 102 NKEELIKAKPDILILAHESQKNSAGKVLKSLKDKGVKVVVVKDAQSIDETDTFKSIGQLT 161
QY 96 ERENKPTFDVSKKDNQPNVHSQLNESHKEDLQREHSQ 135
Db 162 DREKQAKELVDTEKENVEKIINSVPKHKKQEVFNEVSSK 201

RESULT 19
US-09-134-001C-3033
; Sequence 3033, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3033
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3033

Query Match 10.7%; Score 91; DB 2; Length 442;
Best Local Similarity 21.7%; Pred. No. 0.49;
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;

QY 10 KDTGCVSELKPHRVTTI-QNGKE--MSSTIVSEEDFILPVYKGELEK--ELEK----- 54
```

```
Db 90 EDT-EINPVAQPEVNVVTQIEKGKDFEATVTVRPEVKLGDKYKGLIEIKQETDLSDEBLQ 148
QY 55 -----GYQPDGWEISG--PEGKKDAGVWNLNLSKDTFIK 85
Db 149 ESIDHSLSHLAEMVVKEDGAVENGDTVNIDFSG-SVDGEEDFGGQAGYDLEIGSGSIP 207
QY 86 PVFKKIEKKKEENK-----PTFDVS-----KKDNQPNVHSQLNE- 121
Db 208 GFBEQIEGMKTGDEKDVVVVTFPEYHABELAGKEATFKTVKNEIKFKDVPPELNDIEANEL 267
QY 122 -----SHRKEDLQREHSQKSDSTKDVATATVLDKNNISKSITNN 160
Db 268 DSDAENVDEYKENTLRKRLSEQKATEAENT-----EKERAINKATEN 308

RESULT 20
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5157
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5157

Query Match 10.7%; Score 91; DB 2; Length 902;
Best Local Similarity 24.9%; Pred. No. 1.3; Mismatches 31; Indels 24; Gaps 8;
Matches 43; Conservative 31;

QY 1 TTVKEFILNKDTGTVSELKPHRVTTIQNGKMS-STIVSEEDFILPVYKGELEKGYQFD 59
Db 736 TPVSEYRLSNRGGKIK----TATITERNGNIVCITTVTGSEDLMAVVTNAGVI---IRLD 788
QY 60 GWEISGPEGKKDAGY-VINLSKDTFIKPVPKKIEKKEEN-----KPTFDVSKKK 109
Db 789 VHDISQ-NGRAAQGVRLMKLGQGVSTVAKVNEEDDNEERNADEAQAOSTTTTETADVEVV 847
QY 110 DNPQVHNSQLNESHKEDLQREH-HSQKSDSTKDVATATVLDKNNISKSITNN 161
Db 848 D----DQTGPAIHTEGDAEMESVESPENDDRIDRQDPMDRVNEDIESADN 896

RESULT 21
US-08-910-925-3
; Sequence 3, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,925  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0365 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 743 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1684847  
US-08-910-925-3

Query Match 10.7%; Score 90.5; DB 2; Length 743;  
Best Local Similarity 23.8%; Pred. No. 1.1;  
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;  
QY 17 ELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGE-----LEKGYQFGWEI--SGF 66  
DB 281 EARPRR-----QSMKEKEHVVRNEEHKAEQEGKVAQREBELVETGNQHNVDVEIEAGE 335  
QY 67 EGKDGAGVINLSKDTFKIPVKFKIEEKEEENKPTFDVSKKKNPQVNHSHKESHRKE 126  
DB 336 EEEKEIGIVHSDAE-----KEQEEBQKQEMEVKMEEE-----TEVRESEKQQ 378  
QY 127 DLQREHS---QKSDSTKDVT--TVLDKNNISSKSTNN 161  
DB 379 DSQPEVMDVLEWVENVKGVIAQDE/METNRVESVEPSEN 418

RESULT 22  
US-09-949-016-6261  
Sequence 6261, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 6261  
LENGTH: 743  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-6261  
Query Match 10.7%; Score 90.5; DB 2; Length 743;

Best Local Similarity 23.8%; Pred. No. 1.1;  
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;  
QY 17 ELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGE-----LEKGYQFGWEI--SGF 66  
DB 281 EARPRR-----QSMKEKEHVVRNEEHKAEQEGKVAQREBELVETGNQHNVDVEIEAGE 335  
QY 67 EGKDGAGVINLSKDTFKIPVKFKIEEKEEENKPTFDVSKKKNPQVNHSHKESHRKE 126  
DB 336 EEEKEIGIVHSDAE-----KEQEEBQKQEMEVKMEEE-----TEVRESEKQQ 378  
QY 127 DLQREHS---QKSDSTKDVT--TVLDKNNISSKSTNN 161  
DB 379 DSQPEVMDVLEWVENVKGVIAQDE/METNRVESVEPSEN 418

RESULT 23  
US-09-949-016-8288  
Sequence 8288, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 8288  
LENGTH: 758  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-8288

Query Match 10.7%; Score 90.5; DB 2; Length 758;  
Best Local Similarity 28.1%; Pred. No. 1.1;  
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
QY 13 GEVSELKPHRVTVTIQNGKMSSTIVSE--EDFILPVYKGELEKGYQF-----DQWE--- 62  
DB 399 GELOOLSGSQL-----HGKSDSPNVYTEKKGIAILRLRLTELEKRLTPEQQRSDLWRLY 453  
QY 63 -----ISGFEGKQDAG---YVINLSKDTFKIPVKFKIEEKEEENKPTFDVSKKKN 110  
DB 454 VEAKDQNGKQGTGKKKGGRSHRAKNKSKTFGLSV-----KETFDAMKNST 501  
QY 111 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVATATVLDK 150  
DB 502 KEFVRHHEKIKQAKEA--VKENLKKFSDSVKSTFRHPKDTTKNIFDE 547

RESULT 24  
US-09-248-796A-21451  
Sequence 21451, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
SUSCEPTIBILITY TO ANTIFUNGAL AGENTS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13

```
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21451
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21451

Query Match
  10.6%; Score 89.5; DB 2; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.34;
Matches 43; Conservative 26; Mismatches 43; Indels 53; Gaps 9;

QY 10 KDTGEVSEL-KHRVTVT-----IQNGKMSSTIVSEEDP-----43
DB 52 KSTPKTSLRKPVPKPTVVRKWKASKRPPSVTNTPEIKPKSSSPPIISSEDFLEWMD 111
QY 44 ----ILPVYKGELEKGYQFDGWEISGFEKGDAGVYVNLKSDTKFKPVFKKIEKKKEEN 99
DB 112 KLTEVEPIIE-----FNDNYD--NDEDKKE--VVKSKKE-NQNTKGHELPKPKP 160
QY 100 KPTFDVSKKQNPQVNHSQLNSHRKEDLQREHSQKSDTKDVT 144
DB 161 KPSEETSKTKVKPQ-----PRK---QKQKKPLSEETVDTLT 193

RESULT 25
US-09-464-483-4
; Sequence 4, Application US/09464483
; Patent No. 6228617
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6228617el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/464,483
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-464-483-4

Query Match
  10.6%; Score 89.5; DB 2; Length 402;
Best Local Similarity 22.1%; Pred. No. 0.61;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKGELEK-----54
DB 53 DETDIKPVAQPEVSVTIKKGKDFIPEATVTVPEPVKGLDYKGLSIEKQETLSDELQ 112
QY 55 -----GYQFDGWEISG--FEGKGDAGVYVNLKSDTKFKP 86
DB 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG--VDGEEFEGGQAGYDLEIGSGSFI-P 170
QY 87 VPK-----KIEKKKE-----EE--NKPTFDVS-----KKONPQVNHSQLNE- 121
DB 171 GFEEQLEGKMKVDEEKDVVVVTFPEYHABELAGKATFKTKVNEIKFKEVPELTDEIANEL 230

US-09-414-664-4
; Sequence 4, Application US/09414664
; Patent No. 6242249
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6242249el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,664
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-414-664-4

Query Match
  10.6%; Score 89.5; DB 2; Length 402;
Best Local Similarity 22.1%; Pred. No. 0.61;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKGELEK-----54
DB 53 DETDIKPVAQPEVSVTIKKGKDFIPEATVTVPEPVKGLDYKGLSIEKQETLSDELQ 112
QY 55 -----GYQFDGWEISG--FEGKGDAGVYVNLKSDTKFKP 86
DB 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG--VDGEEFEGGQAGYDLEIGSGSFI-P 170
QY 87 VPK-----KIEKKKE-----EE--NKPTFDVS-----KKONPQVNHSQLNE- 121
DB 171 GFEEQLEGKMKVDEEKDVVVVTFPEYHABELAGKATFKTKVNEIKFKEVPELTDEIANEL 230
```

RESULT 30  
US-09-414-664-2  
; Sequence 2, Application US/09414664



```
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Bidhinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-5

Query Match 10.3%; Score 87.5; DB 2; Length 1315;
Best Local Similarity 25.0%; Pred. No. 5;
Matches 41; Conservative 23; Mismatches 73; Indels 27; Gaps 5;

QY 9 NKDTGEVSELKP-HRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQDFGWSISGF- 66
DB 804 NKDGKQDSTKGIQGVTVTKN-----ENGEVLQTTKTDKDGKYQFTGLENGTYK 853
QY 67 -EGKDDAGVNLSDTKFKPVKKEEKEENKPTFDVSKKDNQVNHSQLNESHK 125
DB 854 VEFETPGSYTPT-----QVGSQTDEGDSNGSTSTGVKDKDNDRIDSGFYKPTNKL 905
QY 126 EDLQREHSHQSDSTKD-----VTATVLDKNNISSKSTNNPN 163
DB 906 GDVWVEDTKNGVDKDEKIGSVTVTLKDENDKVLKVTITDEN 949

RESULT 34
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
```

```
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-5

Query Match 10.3%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 7.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSH- 134
DB 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355
QY 135 --QKSDSTKQVTVATVLDKNNISSKSTNNPNK 164
DB 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 35
US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
; US-09-839-996-5

Query Match 10.3%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 7.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSH- 134
DB 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355
```





Db 1296 INTGSAATATTAETAEKSDRQPTETAATSTEDASQHKANTVADNSVANNSSSEPKSERRRSI 1355  
QY 135 --QKSDSTKDVATVLDKNNISSKSTNNPNK 164  
Db 1356 SQOQTSABETTAASTDETTIADNSKRSPNR 1387

RESULT 39

US-09-830-230A-651  
; Sequence 651, Application US/09830230A  
; Patent No. 6902893  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; FILE REFERENCE: PB481US  
; TITLE OF INVENTION: Lyme disease vaccines  
; CURRENT APPLICATION NUMBER: US/09/830,230A  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/057,483  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/053,344  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/053,377  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/050,359  
; PRIOR FILING DATE: 1997-06-20  
; NUMBER OF SEQ ID NOS: 756  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 651  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-830-230A-651

Query Match 10.2%; Score 86.5; DB 2; Length 277;  
Best Local Similarity 20.8%; Pred. No. 0.74; Indels 57; Gaps 4;  
Matches 32; Conservative 26; Mismatches 57; Indels 39; Gaps 4;  
QY 3 VKFEILNKDTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWE 62  
Db 151 IENFFQNDLLFVLTKKNNNTINIMLPNDIQPKDKYILKDKDTIKKG----- 203  
QY 63 ISGFEGKQAGVYINLSKDTTFIKPVFKIEEKEENKPTFVSVKKKDNPNQVNHSQLNES 122  
Db 204 ----TGEK-----YLNPIYR-----FQIKNKQDYHSIDYNKVTIS 234  
QY 123 HRKEDLQREHSQKSDSTKDVTA---TVLDKNNI 153  
Db 235 ERTIYLDLPLHPQVFMNKNFTKILDTITDLNNL 268

RESULT 40

US-08-533-669A-18  
; Sequence 18, Application US/08533669A  
; Patent No. 5834592  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/533,669A  
; FILING DATE: 22-SEP-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-533-669A-18  
Query Match 10.2%; Score 86.5; DB 1; Length 732;  
Best Local Similarity 23.6%; Pred. No. 2.8; Indels 39; Gaps 8;  
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;  
QY 6 FILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65  
Db 170 FTVRTDTGE-----PMGRGTKVILHLKEDQTEYLSERRI-----KEIVKKHSQFIGYPITL 220  
QY 66 F-EKKQDAGVYINLSKDTTFIKPVFKIEEKEENKPTFVSVKKKDNPNQVNHSQLNESHR 124  
Db 221 FVEKERDK----EVSDD-----EAEKEKDKKEEKEKESEDKPEI----- 258  
QY 125 KEDLQREHSQKSD-----STKDVTAATVLDKNNISKST--TNNPN 163  
Db 259 -EDVGSDEEDKDGDKKKKKKKIKKIKYIDKEELNKTPIWTRNPD 302  
Search completed: April 24, 2006, 15:03:39  
Job time : 26.4989 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 109.188 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KLVKDPARTTVTKFILNK.....ATVLKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_21.\*
- 1: Geneseq1980s.\*
  - 2: Geneseq1990s.\*
  - 3: Geneseq2000s.\*
  - 4: Geneseq2001s.\*
  - 5: Geneseq2002s.\*
  - 6: Geneseq2003as.\*
  - 7: Geneseq2003bs.\*
  - 8: Geneseq2004s.\*
  - 9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	773	4	AAB48343 S. pneumo
2	897	100.0	2120	3	AAY81710 Streptoco
3	897	100.0	2140	6	ABU01020 S. pneumo
4	897	100.0	2140	6	ABU45746 Protein e
5	897	100.0	2140	8	ADM92113 S. pneumon
6	897	100.0	2140	8	ADT50099 S. pneumon
7	894	99.7	637	8	ADR94534 Novel S.
8	894	99.7	637	9	AEA58404 Streptoco
9	894	99.7	2138	8	ADK48759 Streptoco
10	615	68.6	117	2	AAW55096 Streptoco
11	615	68.6	117	5	ABP54590 Streptoco
12	615	68.6	117	7	ADC45149 S. pneumo
13	121	13.5	778	5	ABP39023 Staphyloc
14	121	13.5	778	8	ADS06368 Staphyloc
15	119	13.3	746	4	AAG81779 S. epider
16	112.5	12.5	354	9	ADZ72253 Plasmodiu
17	111.5	12.4	707	6	ABU25018 Protein e
18	111.5	12.4	775	6	ABU42797 Protein e
19	110.5	12.3	647	9	ADZ79635 P. falcip
20	110.5	12.3	651	8	ADO19012 Amino aci
21	110.5	12.3	651	8	ADO19010 P. falcip
22	109	12.2	665	3	AB182878 Plasmodiu
23	109	12.2	665	7	ABO23606 Plasmodiu
24	108	12.0	188	9	ADZ79639 P. falcip

25	107.5	12.0	470	8	ADT56185 Plant pol
26	107.5	12.0	484	3	AAG47777 Arabidops
27	105	11.7	1791	8	ADP25441 Plasmodiu
28	104.5	11.6	903	6	ABU24404 Protein e
29	103.5	11.5	169	9	ADZ79634 P. falcip
30	102	11.4	645	9	ADW88441 Staphyloc
31	101.5	11.3	564	4	ABE61977 brosophi
32	100	11.1	1875	6	ABR53560 Protein s
33	100	11.1	1875	7	ADK64380 Disease t
34	100	11.1	1875	8	ADS43855 Bacterial
35	99.5	11.1	635	8	ADS93954 Fibrinoge
36	99.5	11.1	635	8	ADV83292 Streptoco
37	99.5	11.1	643	8	ADV89902 Streptoco
38	99.5	11.1	643	8	ADV81155 Streptoco
39	99.5	11.1	1384	6	ABP55413 Human MDD
40	99.5	11.1	1404	6	ABP55393 Human MDD
41	98.5	11.0	2060	5	AAE20967 Staphyloc
42	98	10.9	441	7	ADS92468 B. lichen
43	97.5	10.9	645	9	ADW88460 Staphyloc
44	97.5	10.9	645	9	ADW88459 Staphyloc
45	97.5	10.9	645	9	ADW88458 Staphyloc

ALIGNMENTS

RESULT 1  
AAB48343  
ID AAB48343 standard; protein; 773 AA.  
XX  
AC AAB48343;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Sp130 polypeptide.  
XX  
KW Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000WO-US015925.  
XX  
PR 10-JUN-1999; 99US-0138453P.  
XX  
(MEDI-) MED IMMUNE INC.  
XX  
PI Adamou JB, Choi GH;  
XX  
WPI; 2001-112197/12.  
DR N-PSDB; AAC84742.  
XX  
New vaccines comprising Sp128 or Sp130 polypeptides, for treating and preventing pneumococcal infections, particularly infections caused by Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or blood infections.

Claim 8; Page 51-54; 54pp; English.

The invention relates to novel immunogenic polypeptides, Sp128 and Sp130 from S. pneumoniae. Vaccines comprising the polypeptides are useful for the treatment and prevention of pneumococcal infections, particularly infections caused by Streptococcus, such as otitis media, nasopharyngeal, bronchial, lung or blood infections. The antigens are used as immunogenic agents to stimulate an immune response. The antisera and antibodies may also be used in diagnosing and treating pneumococcal infections. Recombinant polypeptides serve as a mechanism for stimulating production of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 897; DB 4; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-77;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKEPILNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 60  
 DB 600 KIVVKDFARNTTVKEPILNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 659  
 QY 61 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDD 120  
 DB 660 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDD 719  
 QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 174  
 DB 720 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; protein; 2120 AA.  
 XX  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB002452.  
 XX  
 PR 27-JUL-1998; 98GB-00016336.  
 PR 19-MAR-1999; 99US-0125329P.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWP, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AAZ91806.  
 XX  
 PT Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections.  
 XX  
 PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis  
 XX  
 SQ Sequence 2120 AA;

Query Match 100.0%; Score 897; DB 3; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKEPILNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 60  
 DB 1913 KIVVKDFARNTTVKEPILNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 1972  
 QY 61 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDD 120  
 DB 1973 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDD 2032  
 QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 174  
 DB 2033 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 1; SEQ ID NO 1180; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 897; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 60  
 DB 1933 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKKEENKPTFDVSKKGD 120  
 DB 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKKEENKPTFDVSKKGD 2052

QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.  
 AC ABU45746;  
 XX  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #31273.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 XX Streptococcus pneumoniae.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA49616.  
 DR

XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 73670; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 897; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 60  
 DB 1933 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKKEENKPTFDVSKKGD 120  
 DB 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKKEENKPTFDVSKKGD 2052

QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.  
 AC ADM92113;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE S pneumoniae antigenic protein sequence SeqID310.  
 XX  
 KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 KW antigenic.  
 XX  
 XX Streptococcus pneumoniae.  
 OS

PN WO2004020609-A2.  
 XX  
 PD 11-MAR-2004.  
 XX  
 PF 02-SEP-2003; 2003WO-US027401.  
 XX  
 PR 30-AUG-2002; 2002US-0407082P.  
 XX  
 PA (TUPT ) UNIV TUPTS.  
 XX  
 PI Camilli A, Hava DL;  
 XX  
 DR MPI; 2004-239189/22.  
 DR N-PSDB; ADM91876.  
 XX  
 PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
 PT diagnosing, treating and preventing active infections of Streptococcus  
 PT pneumoniae.  
 XX  
 PS Claim 27; SEQ ID NO 310; 123pp; English.  
 XX  
 CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
 CC acid molecules and the antigenic polypeptides encoded by them. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
 CC compositions and methods disclosed are useful for treating Streptococcus  
 CC pneumoniae infection. The present sequence is that of an S pneumoniae  
 CC protein of the invention.  
 XX  
 SQ Sequence 2140 AA;  
 XX  
 Query Match 100.0%; Score 897; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60  
 DB 1933 KIVVDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1992  
 QY 61 ELEKGYPDQWBEISGFEKGKQAGYVNLKDTFKVPFKIEEKKEENKPTFDVSKKKD 120  
 DB 1993 ELEKGYPDQWBEISGFEKGKQAGYVNLKDTFKVPFKIEEKKEENKPTFDVSKKKD 2052  
 QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106  
 RESULT 6  
 ID ADT50099  
 XX ADT50099 standard; protein; 2140 AA.  
 AC ADT50099;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
 XX  
 KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
 KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
 KW sepsis; meningitis.  
 XX  
 OS Streptococcus pneumoniae TIGR4.  
 XX  
 PN WO2004092209-A2.  
 XX  
 PD 28-OCT-2004.  
 XX  
 PF 15-APR-2004; 2004WO-EP003984.  
 XX  
 PR 15-APR-2003; 2003EP-00450087.  
 XX  
 PA (INTE-) INTERCELL AG.

XX  
 PI Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
 XX  
 DR MPI; 2004-758335/74.  
 DR N-PSDB; ADT49955.  
 XX  
 PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
 PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
 PT treating S. pneumoniae infections.  
 XX  
 PS Disclosure; SEQ ID NO 177; 191pp; English.  
 XX  
 CC This invention relates to novel nucleic acids encoding hyperimmune serum  
 CC reactive antigens, or fragments derived thereof. Specifically, it refers  
 CC to antigens selected from peptides and serum reactive epitopes that can  
 CC be used in pharmaceutical compositions that exhibit antibacterial  
 CC activity. The present invention describes a composition (including the  
 CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
 CC that is useful for manufacturing a medicament such as a vaccine, which  
 CC can be used to treat or prevent bacterial infections, particularly S.  
 CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
 CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
 CC be used for isolating, purifying and/ or identifying an interaction  
 CC partner of the hyperimmune serum reactive antigen, as well as for  
 CC manufacturing a functional nucleic acid selected from aptamers and  
 CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
 CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
 CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
 CC of the invention.  
 XX  
 SQ Sequence 2140 AA;  
 XX  
 Query Match 100.0%; Score 897; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60  
 DB 1933 KIVVDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1992  
 QY 61 ELEKGYPDQWBEISGFEKGKQAGYVNLKDTFKVPFKIEEKKEENKPTFDVSKKKD 120  
 DB 1993 ELEKGYPDQWBEISGFEKGKQAGYVNLKDTFKVPFKIEEKKEENKPTFDVSKKKD 2052  
 QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106  
 RESULT 7  
 ID ADR94534  
 XX ADR94534 standard; protein; 637 AA.  
 AC ADR94534;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
 XX  
 KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US6800744-B1.  
 XX  
 PD 05-OCT-2004.  
 XX  
 PF 30-JUN-1998; 98US-00107433.  
 XX  
 PR 02-JUL-1997; 97US-0051553P.  
 PR 12-MAY-1998; 98US-0085131P.  
 XX  
 XX

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2004-697205/68.  
XX N-PSDB; ADR91931.  
XX  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequences.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=680074481.  
XX  
XX Sequence 637 AA;  
SQ  
Query Match 99.7%; Score 894; DB 8; Length 637;  
Best Local Similarity 99.4%; Pred. No. 9.8e-77;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 430 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 489  
QY 61 ELEKGYQPDGWEISGFEKGKDGAGYVNLKDTFFIKPVFKKIEKKEENKPTFDVSKKD 120  
DB 490 ELEKGYQPDGWEISGFEKGKDGAGYVNLKDTFFIKPVFKKIEKKEENKPTFDVSKKD 549  
QY 121 NPQVNHSQLNESHKEDLQREHDSQKSDTKDVTATVLDKNNISSKSTTNPNK 174  
DB 550 NPQVNHSQLNESHKEDLQREHDSQKSDTKDVTATVLDKNNISSKSTTNPNK 603  
RESULT 8  
AEA58404  
ID AEA58404 standard; protein; 637 AA.  
XX  
XX AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
XX  
XX Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
DE  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
KW vaccine.  
XX  
XX Streptococcus pneumoniae.  
OS  
XX  
XX US2005136404-A1.  
FN  
XX

PD 23-JUN-2005.  
XX  
PF 10-JUL-2003; 2003US-00617320.  
XX  
XX 02-JUL-1997; 9TUS-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
XX (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
XX New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
XX Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - the sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
XX Sequence 637 AA;  
SQ  
Query Match 99.7%; Score 894; DB 9; Length 637;  
Best Local Similarity 99.4%; Pred. No. 9.8e-77;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 430 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 489  
QY 61 ELEKGYQPDGWEISGFEKGKDGAGYVNLKDTFFIKPVFKKIEKKEENKPTFDVSKKD 120  
DB 490 ELEKGYQPDGWEISGFEKGKDGAGYVNLKDTFFIKPVFKKIEKKEENKPTFDVSKKD 549  
QY 121 NPQVNHSQLNESHKEDLQREHDSQKSDTKDVTATVLDKNNISSKSTTNPNK 174  
DB 550 NPQVNHSQLNESHKEDLQREHDSQKSDTKDVTATVLDKNNISSKSTTNPNK 603  
RESULT 9  
ADK48759  
ID ADK48759 standard; protein; 2138 AA.

```
XX ADK48759;
AC
XX
XX 20-MAY-2004 (first entry)
DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
XX Streptococcus pneumoniae.
OS
XX US6699703-B1.
PN
XX
XX 02-MAR-2004.
PD
XX
XX 26-MAY-2000; 2000US-00583110.
PF
XX
XX 02-JUL-1997; 97US-0051553P.
PR
XX 12-MAY-1998; 98US-0085131P.
PR
XX 30-JUN-1998; 98US-00107433.
PR
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
XX WPI; 2004-212399/20.
DR
XX N-PSDB; ADK46098.
DR
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
XX Disclosure; SEQ ID NO 5274; 301pp; English.
XX
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
SQ
Query Match 99.7%; Score 894; DB 8; Length 2138;
Best Local Similarity 99.4%; Pred. No. 5.2e-76;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KIVVKDPARNTTVKEPILNKDTGEVSELKPHRVTVTTIQGKMSSTIVSEEDFILPVYKG 60
Db 1931 KIVVKDPARNTTVKEPILNKDTGEVSELKPHRVTVTTIQGKMSSTIVSEEDFILPVYKG 1990
QY 61 ELEKGQFDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSKKD 120
Db 1991 ELEKGQFDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSKKD 2050
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 174
Db 2051 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX
XX AAW55096;
AC
XX
XX 02-OCT-1998 (first entry)
DT
XX Streptococcus pneumoniae SP043 protein.
DE
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
OS
XX WO9818930-A2.
PN
XX
XX 07-MAY-1998.
PD
XX
XX 30-OCT-1997; 97WO-US019422.
PF
XX
XX 31-OCT-1996; 96US-0029960P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX WPI; 1998-272224/24.
DR
XX N-PSDB; AAV27357.
DR
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX Sequence 117 AA;
SQ
Query Match 68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.9e-51;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 YKGELEKGYQFDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 117
Db 1 YKGELEKGYQFDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 60
QY 118 KDNPPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 174
Db 61 KDNPPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
XX AC ABP54590;
AC
XX
XX 04-SEP-2002 (first entry)
DT
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
DE
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
OS
XX US2002061545-A1.
PN
```



```
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2002-479261/51.
XX N-PSDB; ABQ84825.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX and for preventing or attenuating disease caused by Streptococcus
XX infection.
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX pneumoniae antigens have antibacterial activity and can be used in
XX vaccines. The S. pneumoniae antigens can also be used to prevent or
XX attenuate a Streptococcal infection in an animal. The polynucleotides
XX encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX of S. pneumoniae ORFs (open reading frames) which are used in an example
XX from the present invention
XX
XX Sequence 117 AA;
XX
XX Query Match 68.6%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5,9e-51;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 117
XX Db 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 60
XX
XX QY 118 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
XX Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 12
XX ADC45149
XX ID ADC45149 standard; protein; 117 AA.
XX
XX AC ADC45149;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE S. pneumoniae antigenic protein SP043.
XX
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
XX OS Streptococcus pneumoniae.
XX
XX XX US6573082-B1.
XX
XX PD 03-JUN-2003.
XX
XX PF 28-MAR-2000; 2000US-00536784.
XX
XX PR 31-OCT-1996; 96US-0029960P.
```

```
PR 30-OCT-1997; 97US-00961083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2003-764574/72.
XX N-PSDB; ADC45148.
XX
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX useful for producing vaccines for prevention or attenuation of infection
XX by Streptococcus pneumoniae.
XX
XX Example 1; SEQ ID NO 68; 58pp; English.
XX
XX The invention relates to an isolated polynucleotide consisting of a
XX Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX antigens. Also included are making a recombinant vector by inserting the
XX nucleic acid into a vector, an isolated polynucleotide consisting of at
XX least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX acids are useful as DNA vaccine against Streptococcus pneumoniae
XX infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX antigen nucleic acids are useful as probes for use in diagnostic methods
XX for detecting S. pneumoniae gene expression. The present sequence
XX represents an S. pneumoniae antigenic protein.
XX
XX Sequence 117 AA;
XX
XX Query Match 68.6%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5,9e-51;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 117
XX Db 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 60
XX
XX QY 118 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
XX Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 13
XX ABP39023
XX ID ABP39023 standard; protein; 778 AA.
XX
XX AC ABP39023;
XX
XX DT 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.
XX
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX OS Staphylococcus epidermidis.
XX
XX PN US6380370-B1.
XX
XX PD 30-APR-2002.
XX
XX PF 13-AUG-1998; 98US-00134001.
XX
XX PR 14-AUG-1997; 97US-0055779P.
XX PR 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
```

```

DR N-PSDB; ABN91568.
XX
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
XX Disclosure; SEQ ID NO 3868; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP5124 to ABP3960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ Sequence 778 AA;
Query Match 13.5%; Score 121; DB 5; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.022;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;
QY 6 DPAR--NTTVKBFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK---- 59
DB 601 DFSKVPNVQGDVEQKAEDSVNAQSLKP-----ITIGNGKIQKQSVKSGTKVLPHSKVLM 656
QY 60 --GELEKGYQFDGW---EISGFE-----GKKDAGYVIN--LSKDTFIKPVFKKIEKK 105
DB 657 TDGELTMP-DMTGWTKEVDLAFEDLTIKVSTKNGGFVTVNQISKGQIIK----- 705
QY 106 EBEENKPTFDVS-----KKKDNPNQVNHSQLNESHKEDLQREHHSOKSDSKDVTATVLDKN 161
DB 706 ---NKDKIEVLSAEDTDDQEKTDDESDNKSKKDADEHSTSSSTKN-----DKS 756
QY 162 NISSKSTNN 171
DB 757 NADSKNDSDD 766
RESULT 14
AD506368
ID ADS06368 standard; protein; 778 AA.
XX
AC ADS06368;
XX
XX 04-NOV-2004 (first entry)
XX
XX Staphylococcus epidermis polypeptide seqid 5663.
XX
XX antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
XX recombinant expression vector; infection; computer readable medium;
XX computer based system.
XX
XX Staphylococcus epidermidis.
XX
XX US2004147734-A1.
XX
XX 29-JUL-2004.
XX
XX 01-DEC-2003; 2003US-00724972.
XX
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX
XX Doucette-Stamm L, Bush D;
XX
XX WPI; 2004-580138/56.
XX

```

```

DR N-PSDB; ADS02596.
XX
XX
PT New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
XX Claim 17; SEQ ID NO 5663; 741pp; English.
XX
XX The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.
XX
SQ Sequence 778 AA;
Query Match 13.5%; Score 121; DB 8; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.022;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;
QY 6 DPAR--NTTVKBFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK---- 59
DB 601 DFSKVPNVQGDVEQKAEDSVNAQSLKP-----ITIGNGKIQKQSVKSGTKVLPHSKVLM 656
QY 60 --GELEKGYQFDGW---EISGFE-----GKKDAGYVIN--LSKDTFIKPVFKKIEKK 105
DB 657 TDGELTMP-DMTGWTKEVDLAFEDLTIKVSTKNGGFVTVNQISKGQIIK----- 705
QY 106 EBEENKPTFDVS-----KKKDNPNQVNHSQLNESHKEDLQREHHSOKSDSKDVTATVLDKN 161
DB 706 ---NKDKIEVLSAEDTDDQEKTDDESDNKSKKDADEHSTSSSTKN-----DKS 756
QY 162 NISSKSTNN 171
DB 757 NADSKNDSDD 766
RESULT 15
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX
AC AAG81779;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX endocarditis.
XX
XX Staphylococcus epidermidis.
XX

```

PN W0200134809-A2.  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US030782.  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.  
XX N-PSDB; AAH52629.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX Claim 18; Page 208; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX Sequence 746 AA;  
SQ  
Query Match 13.3%; Score 119; DB 4; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.032;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
QY 20 KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDG-- 71  
DB 585 EDSVNAQSLKP-----ITINGKQIKQOQSVKSGTKVLPHSKVLMLTGELTMP-DMTGWTK 639  
QY 72 -EISGFE-----GKKDAGYVIN--LSKDTFTKPVFKIKIEKKEBENKPTFDVS---K 117  
DB 640 EDVLAPEDTLKLKVSFKNGFVTNQISIKQIILK-----NKKIEVLSAED 686  
QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTVATVLDKNNISSKSTNN 171  
DB 687 TDDQKQETDSDSDNKSXKQKADHDHSNTSSSTKN-----DKSNADSKNDSDD 734  
RESULT 16  
ADZ72253  
ID ADZ72253 standard; protein; 354 AA.  
XX ADZ72253;  
XX 14-JUL-2005 (first entry)  
XX Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.  
DE Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.  
XX

OS Plasmodium falciparum.  
XX EP1526178-A1.  
XX 27-APR-2005.  
XX 24-OCT-2003; 2003EP-00292673.  
XX 24-OCT-2003; 2003EP-00292673.  
XX (INSP ) INST PASTEUR.  
XX Drulhe P;  
XX WPI; 2005-323987/34.  
XX N-PSDB; ADZ72252.  
XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
PT falciparum, which encode proteins useful for preparing vaccine  
PT compositions against malaria.  
XX Disclosure; SEQ ID NO 2; 137pp; English.  
XX The present invention relates to the protection against malaria. More  
CC particularly, the invention pertains to a family of MSP-3 (merozoite  
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
CC falciparum, highly conserved in P. falciparum strains, simultaneously  
CC expressed in P. falciparum at the erythrocytic stages and encoding  
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
CC their N-terminal extremity and which are located at the merozoite  
CC surface. The characterization of this gene family enables the definition  
CC of immunogenic and vaccine compositions against P. falciparum. The  
CC present sequence is the P. falciparum MSP-3-1 protein.  
XX Sequence 354 AA;  
SQ  
Query Match 12.5%; Score 112.5; DB 9; Length 354;  
Best Local Similarity 22.4%; Pred. No. 0.048;  
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;  
QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF 69  
DB 153 STTKTEYAEKAKNAYEKAKNAYQKANAQAVLKAKBASS-----YDYIL----- 194  
QY 69 DGWEISGR--EKKDAG-----YVINLSKDTFTKPVFKIKIEKKEB-----E 108  
DB 195 -GWFGGGVPEKKKKNMLSHLYVSKKKNISKENDVDVLDKKEBAAETEEHEEKEKNE 253  
QY 109 NKPTFDVSKKNQPNVNHSQLNESHKEDLQREHSQKSDSTKQVTVATVLDKNNISSKST 168  
DB 254 BETEISEIDEEEBEKEBENDKKKEQKEQSNENNNDQKKDMEA-----QNLISKQ 308  
QY 169 TNN 171  
DB 309 NNN 311  
RESULT 17  
ABU25018  
ID ABU25018 standard; protein; 707 AA.  
XX ABU25018;  
XX 19-JUN-2003 (first entry)  
XX Protein encoded by Prokaryotic essential gene #10545.  
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
KW Clostridium difficile.  
XX WO200277183-A2.  
PN

```
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA28888.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 52942; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 707 AA;
Query Match 12.4%; Score 111.5; DB 6; Length 707;
Best Local Similarity 26.6%; Pred. No. 0.15;
Matches 47; Conservative 28; Mismatches 63; Indels 39; Gaps 7;
QY 2 IVVKDFARNVTYKFEI--LNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK 59
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 ISIEDDAEE-GVKEIDSNQDGDVVEDKD-----TTDKEDYS---NKEDIISPENK 532
QY 60 GELEKGYQFGWEISGPEGKDGAGYVNLKSTFIKPVFKIEEKEEN--KPTDVS 117
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 KSKKAKLFG-----FIKKNEEVSQEEENLNDISPDILDK 569
QY 118 KQNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNISSSKSTNNPNK 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 570 PVENNQVKSEIEQNELKE-IKQEPFSQHIEBERSVKIEPINNLDKVSNNESK 625
RESULT 18
ABU42797
ID ABU42797 standard; protein; 775 AA.
XX AC ABU42797;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #28324.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Staphylococcus epidermidis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA46667.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 70721; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
```

```
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 775 AA;

Query Match 12.4%; Score 111.5; DB 6; Length 775;
Best Local Similarity 24.9%; Pred. No. 0.18;
Matches 42; Conservative 28; Mismatches 33; Gaps 8;

QY 20 KDTGVESELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GHELEKGYQPDGW-- 71
DB 614 EDSVNAQSLKP-----ITIGNGKQIKQOSVKSQTKVLPKSHKVLMTDGLTMP-DMTGWTK 668
QY 72 -EISGPE-----GKDDAGYVIN--LSKDTFTKPVFK-----KISEKKEENKPTF 113
DB 669 EDVLAFEDLTAKIKVSTKNGGFTVNTQISKGQIIKNDKIEVLSABEDTDDQKTDSDSS 728
QY 114 DVSKKDNQPNVHSHRQEDLQREHSHQSKSDSTKDVTAIVLDKNN 162
DB 729 DKSKDKQVDEDSNASSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 19
ADZ79635
ID ADZ79635 standard; protein; 647 AA.
XX AC ADZ79635;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum GLURP-MSP3 fusion protein.
XX KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;
XX KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP ) INST PASTEUR.
XX PI Druilhe P;
XX WPI; 2005-355821/36.
XX N-PSDB; ADZ79636.

Chimeric molecule useful for preparing vaccine composition against
malaria, comprises glutamate-rich protein GLURP and Merozoite surface
protein 3 MSP3 moieties, and raises antibodies against moieties in mice
immunized with molecule.

Disclosure; SEQ ID NO 3; 79pp; English.

The invention relates to a chimeric molecule that comprises a glutamate-
rich protein (GLURP) moiety consisting of a polypeptide fragment (amino
acid residues 25-514) of GLURP (given as SEQ ID No.1) and a Merozoite
surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380
of MSP3 (given as SEQ ID No.2), wherein the chimeric molecule raises
antibodies against both polypeptides in mice immunized with it. Also
described are: (i) a conjugate comprising the chimeric molecule of the
invention bound to a solid support, (ii) an immunogenic composition
comprising the chimeric molecule, the conjugate described above, or a
mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine
against malaria comprising the chimeric molecule, the conjugate described
above, or a mixture of GLURP and MSP3 antigens as an immunogen, in
association with a suitable vehicle, (iv) use of purified and/or
```

```
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a
CC medicament against malaria, and (v) a medicament for passive
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP
CC and MSP3 antigens are useful for the preparation of a vaccine composition
CC against malaria. This sequence represents Plasmodium falciparum GLURP(27-
XX 500)-MSP3(212-380) fusion protein.
XX SQ Sequence 647 AA;

Query Match 12.3%; Score 110.5; DB 9; Length 647;
Best Local Similarity 22.8%; Pred. No. 0.17;
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;

QY 14 KEFLNKDTGKRVSELSKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63
DB 418 EEAVSEKNAHETVE---HEETVQSSENPKEADNGVSNQNNELNENEFV-----ESE 468
QY 64 K-----GYQPD-GWEISGF--EGKKDAG-----YVINLSKDTFTKPVVKIE 102
DB 469 KSEHEARSKAKEASSYDYTLGWFGGVPHEKKEENMLSHLYVSSKDKENISKENDDVLD 528
QY 103 EKKERENKPTPDVSKKKNPQVNHSLN-----ESHRKEDLQREHSHQSKSDS 149
DB 529 E-KEEAEETEEELKEENETESIESEDEEEEEEEKEENKKEKQEKQSQSNENNDO 587
QY 150 TKDVTATVLDKNNISKSTNN 171
DB 588 KKDEA-----QNLISKQNNN 604

RESULT 20
AD019012
ID AD019012 standard; protein; 651 AA.
XX AC AD019012;
XX DT 12-AUG-2004 (first entry)
XX DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;
XX KW antimalarial; immunostimulant.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO2004043488-A1.
XX PD 27-MAY-2004.
XX PF 06-NOV-2003; 2003WO-DK000759.
XX PR 12-NOV-2002; 2002DK-00001741.
XX PR 11-SEP-2003; 2003DK-00001307.
XX PA (STAT-) STATENS SERUM INST.
XX PI Theisen M, Jepsen S;
XX WPI; 2004-411650/38.

New antigen based vaccine comprising a fusion protein derived from
Plasmodium falciparum Glutamate-rich protein, useful in treating or
preventing malaria.

Disclosure; Fig 2C; 52pp; English.

The present invention relates to a fusion protein comprising Plasmodium
falciparum glutamate-rich protein (GLURP) coupled to P. falciparum
merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is
useful as an antigen based vaccine against malaria. Also disclosed is the
```



CC of infection with *P. falciparum*. Furthermore, (1) (especially when they  
 CC are rifins or secreted or membrane proteins) can aid the identification  
 CC of drugs to treat or prevent *P. falciparum* infection, or they can be used  
 CC to identify drug resistance in *P. falciparum*. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification

XX SQ Sequence 665 AA;

Query Match 12.2%; Score 109; DB 3; Length 665;  
 Best Local Similarity 26.1%; Pred. No. 0.25; Indels 58; Gaps 11;  
 Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;  
 QY 5 KDFARNTTVKGFILNKDTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELE 63  
 DB 114 KDDNNNNNGTKQIEBKKNKINKSDL--HQNELNLQSGK-----NEQDI-----NKNE 158  
 QY 64 KGYQPDGWEI--SGFEGKKDAGYVNLKDTPIKVPFKKIEKKK-----EENKP 111  
 DB 159 KGKQ----DISNSNAENKDD-----VKEGVKELEBKKEKISDDHKVVENKK 202  
 QY 112 TFD----VSKKKDNPQVNHSQLNESHKEDLQR--EHSQKSDSTKDVTTATVLDKNNISK 166  
 DB 203 SDDHKVVENKSGDDHKVVENKSDDHKIEVKVKEHEDEDEE-----DKKEKKS 253  
 QY 167 STTNPNK 174  
 DB 254 NKNKDKENK 261

RESULT 23

ABO23606  
 ID ABO23606 standard; protein; 665 AA.

XX AC ABO23606;

XX DT 04-SEP-2003 (first entry)

XX DE Plasmodium falciparum outlier protein #3.

XX KW Candidate protein identification; pathogen; anti-infective;  
 XX KW outlier protein; virulence protein; antigen; drug target protein;  
 XX KW pathogenic organism; antimicrobial.

XX OS Plasmodium falciparum.

XX PN US2003039963-A1.

XX PD 27-FEB-2003.

XX PF 30-MAR-2001; 2001US-00820843.

XX PR 30-MAR-2001; 2001US-00820843.

XX PA (BRAH/) BRAHMACHARI S K.

XX PA (RAMA/) RAMACHANDRAN S.

XX PA (NAND/) NANDI T.

XX PA (BHIM/) BHIMARAO C.

XX PI Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;

XX DR WPI; 2003-492159/46.

XX PT Identifying candidate proteins useful as anti-infectives involves  
 PT matching outlier protein sequences with protein sequences in databases.

XX PS

XX Example 7; Page 91-93; 117pp; English.

CC The present invention relates to a method for identifying candidate  
 CC proteins in pathogens useful as anti-infectives. The invention discloses  
 CC a computational method which involves the calculation of several sequence  
 CC attributes and their subsequent analysis results in the identification  
 CC of outlier proteins in different pathogens. The method is useful for the  
 CC identification of outlier proteins (e.g. virulence proteins, antigens or  
 CC proteins used as drug targets) in pathogenic organisms. The method of the  
 CC invention provides reproducible results as it does not depend on the  
 CC variable biochemical characterisation of proteins. ABO23500-ABO23617  
 CC represent outlier proteins identified from different pathogenic organisms

XX SQ Sequence 665 AA;

Query Match 12.2%; Score 109; DB 7; Length 665;  
 Best Local Similarity 26.1%; Pred. No. 0.25; Indels 58; Gaps 11;  
 Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;  
 QY 5 KDFARNTTVKGFILNKDTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELE 63  
 DB 114 KDDNNNNNGTKQIEBKKNKINKSDL--HQNELNLQSGK-----NEQDI-----NKNE 158  
 QY 64 KGYQPDGWEI--SGFEGKKDAGYVNLKDTPIKVPFKKIEKKK-----EENKP 111  
 DB 159 KGKQ----DISNSNAENKDD-----VKEGVKELEBKKEKISDDHKVVENKK 202  
 QY 112 TFD----VSKKKDNPQVNHSQLNESHKEDLQR--EHSQKSDSTKDVTTATVLDKNNISK 166  
 DB 203 SDDHKVVENKSGDDHKVVENKSDDHKIEVKVKEHEDEDEE-----DKKEKKS 253  
 QY 167 STTNPNK 174  
 DB 254 NKNKDKENK 261

RESULT 24

ADZ79639  
 ID ADZ79639 standard; protein; 188 AA.

XX AC ADZ79639;

XX DT 14-JUL-2005 (first entry)

XX DE *P. falciparum* merozoite surface protein 3, MSP3a to MSP3f fragment.

XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 XX KW immunotherapy; malaria; antimalarial; vaccine.

XX OS Plasmodium falciparum.

XX PN WO2005040206-A1.

XX PD 06-MAY-2005.

XX PF 22-OCT-2004; 2004WO-EP012910.

XX PR 24-OCT-2003; 2003US-00691672.

XX PA (INSP ) INST PASTEUR.

XX PI Druilhe P;

XX DR WPI; 2005-355821/36.

XX PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GIURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

XX PS Disclosure; SEQ ID NO 7; 79pp; English.

XX XX



CC The invention relates to a chimeric molecule that comprises a glutamate-rich protein (GLURP) moiety consisting of a polypeptide fragment (amino acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380 of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises antibodies against both polypeptides in mice immunized with it. Also described are: (i) a conjugate comprising the chimeric molecule of the invention bound to a solid support, (ii) an immunogenic composition comprising the chimeric molecule, the conjugate described above, or a mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine against malaria comprising the chimeric molecule, the conjugate described above, or a mixture of GLURP and MSP3 antigens as an immunogen, in association with a suitable vehicle, (iv) use of purified and/or recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a medicament against malaria, and (v) a medicament for passive immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP antibodies. The chimeric molecule of the invention or a mixture of GLURP and MSP3 antigens are useful for the preparation of a vaccine composition against malaria. This sequence represents Plasmodium falciparum MSP3a to MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the Sequence Listing is not mentioned elsewhere in the specification.

XX Sequence 188 AA;

Query Match 12.0%; Score 108; DB 9; Length 188;  
Best Local Similarity 23.2%; Pred. No. 0.054;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 37 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINUS 89  
DB 15 VLKAKEASS-----YDYL-----GWFGGVPKHKEENMLSHLVSSKD 55  
QY 90 KDTFIPKVPKKEEKEE-----ENKPTFVSKKKNQVNVHSQLNESHKXE 136  
DB 56 KENISKENDVLDKEEAEETEEELKEKNEEETSEISEDEEEEEEEKEEENDKKK 115  
QY 137 DLQREBHSOKSDSTKDVATVLDKNNISKSTTN 171  
DB 116 EQEKEQSNENNQKKDMEA-----QNLISKNNQN 145

RESULT 25

ID ADT56185  
ADT56185 standard; protein; 470 AA.

XX ADT56185;

DT 13-JAN-2005 (first entry)

DE Plant polypeptide, SEQ ID 6262.

XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.

OS Viridiplantae.

XX US2004216190-A1.

XX 28-OCT-2004.

XX 18-DEC-2003; 2003US-00739930.

XX 28-APR-2003; 2003US-00424599.

PR 28-APR-2003; 2003US-00425115.

XX (KOVA/) KOVALIC D K.

XX Kovalic DK;

XX WPI; 2004-757369/74.

XX New recombinant DNA constructs useful in the field of biochemistry and genetics, and in particular for producing transgenic plants with improved biological characteristics.

XX Claim 2; SEQ ID NO 6262; 14pp; English.

XX The invention relates a recombinant DNA construct comprising a polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO: 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean, Arabidopsis, wheat and rape but the specification does not indicate which sequences is derived from which organism. Also included is a method of producing a plant having an improved property, comprising transforming a plant with a recombinant DNA construct comprising a promoter region functional in a plant cell operably joined to a polynucleotide encoding a polypeptide associated with the property, and growing the transformed plant. The property is selected from improving plant cold tolerance, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for improving plant drought tolerance, for providing increased resistance to plant disease, for galactomannan production, for production of plant growth regulators, for improving plant heat tolerance, for improving plant tolerance to herbicides, for increasing the rate of homologous recombination in plants, for lignin production, for improving plant tolerance to extreme osmotic conditions, for improving plant tolerance to pathogens or pests, for yield improvement by modification of photosynthesis, for modifying seed oil yield and/or content, for modifying seed protein yield and/or content, for yield improvement by modification of carbohydrate, nitrogen or phosphorus use and/or uptake and for yield improvement by providing improved plant growth and development under at least one stress condition. The polynucleotide may also encode a plant transcription factor. The methods and compositions of the present invention are useful in the field of biochemistry and genetics, in particular for producing transgenic plants with improved biological characteristics such as increased yield, improved nitrogen flow, increasing plant tolerance to cold or heat, improving plant tolerance to extreme osmotic and drought conditions, and improving plant tolerance to plant pests or pathogens. They can also be used in physical arrays of molecules, plant breeding markers, computer-based storage and analysis systems. The present sequence is one of the 5544 plant protein sequences of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DccID=20040216190.

XX Sequence 470 AA;

Query Match 12.0%; Score 107.5; DB 8; Length 470;  
Best Local Similarity 20.1%; Pred. No. 0.21;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 29 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF-----GKK 80

DB 82 RENRVTTDVQNNNGESK-----YVQDLARRIRYDE-EATGSQSAQRIDHPNQK 129

QY 81 DAGYVINSLSKTFIPKVPKKEEKEENKPTFDVSKKKN----- 121

DB 130 NVGITEKAFENSPIETSHRVDNRIINNQNFTAAKSSSENAVSRVSGADHKRAEVMGK 189

QY 122 PQVNHSQLNE-----SHRKEDLQFEHHSQKSDSTKDVATVLDKNNISKSTTNPNK 174

DB 190 PMENRDQVRQTESAEKSHRKENVTISEKPRQEGVKYKTEAKDKRNKKEKTEKTESINK 248

RESULT 26

AAG47777

ID AAG47777 standard; protein; 484 AA.

XX AAG47777;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 60255.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; Genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EF1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-01334256P.  
PR 14-MAY-1999; 99US-0133418P.  
PR 14-MAY-1999; 99US-0133421P.  
PR 14-MAY-1999; 99US-0133422P.  
PR 14-MAY-1999; 99US-01334370P.  
PR 18-MAY-1999; 99US-01334768P.  
PR 19-MAY-1999; 99US-01334941P.  
PR 20-MAY-1999; 99US-01335124P.  
PR 21-MAY-1999; 99US-01335353P.  
PR 24-MAY-1999; 99US-0133629P.  
PR 25-MAY-1999; 99US-01336021P.  
PR 27-MAY-1999; 99US-01336392P.  
PR 28-MAY-1999; 99US-01336782P.  
PR 01-JUN-1999; 99US-01337222P.  
PR 03-JUN-1999; 99US-01337528P.  
PR 04-JUN-1999; 99US-01337502P.  
PR 07-JUN-1999; 99US-01337724P.  
PR 08-JUN-1999; 99US-01338094P.  
PR 10-JUN-1999; 99US-01338540P.  
PR 10-JUN-1999; 99US-01338847P.  
PR 14-JUN-1999; 99US-01339119P.  
PR 16-JUN-1999; 99US-01339452P.  
PR 16-JUN-1999; 99US-01339453P.  
PR 17-JUN-1999; 99US-01339492P.  
PR 18-JUN-1999; 99US-01339454P.  
PR 18-JUN-1999; 99US-01339455P.  
PR 18-JUN-1999; 99US-01339456P.  
PR 18-JUN-1999; 99US-01339457P.  
PR 18-JUN-1999; 99US-01339458P.  
PR 18-JUN-1999; 99US-01339459P.  
PR 18-JUN-1999; 99US-01339460P.  
PR 18-JUN-1999; 99US-01339461P.  
PR 18-JUN-1999; 99US-01339462P.  
PR 18-JUN-1999; 99US-01339463P.  
PR 18-JUN-1999; 99US-01339750P.  
PR 21-JUN-1999; 99US-01339763P.  
PR 21-JUN-1999; 99US-01339817P.  
PR 22-JUN-1999; 99US-01339899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149829P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.



QY 9 RNTTVK--EFILNKTGEVSELKP-----HRVTVTIQNG-----KEMSTTVSE 50  
Db 614 RNTKYKSDYLLDNKGESKPKRNTSYVLESPLHLIGDIVDNINIKRKKKKKIKTIVSD 673  
QY 51 EDPILPVYKGELEKGYQPDGWEISG---PEGKDGAGYVNLKSDTPFKPVPKKIEKKKEE 107  
Db 674 DMFTSPVNIKEVYNEQERKEIVGNLSYDKTKIPFPFKFTKEGRKK--KKIEKKKK 731  
QY 108 ENK-----PTF-----DVSKKDNPNQVNSQL----- 129  
Db 732 EKKEKNNFLYNDYSYSPKYGDNNENNFVIKYIRERKDFOKCPDHPNPFNSKFLNYYN 791  
QY 130 ---NESHK---EDLOREHSQKSDTKD-VTATVLDKNNISKSTTNNPNK 174  
Db 792 PMKNKNNKNNKNNKNNVERNYPNTYSSKDGVSYNFLSDSLFSSDNEYSSDNE 843

RESULT 28

ABU24404  
ID ABU24404 standard; protein; 903 AA.  
AC ABU24404;

19-JUN-2003 (first entry)

Protein encoded by Prokaryotic essential gene #9931.

Antisense; prokaryotic essential gene; cell proliferation; drug design.  
Clostridium botulinum.

WO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(BLIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.

N-PSDB; ACA28274.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 52328; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway; (8) required for proliferation, or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 903 AA;

Query Match 11.6%; Score 104.5; DB 6; Length 903;  
Best Local Similarity 21.9%; Pred. No. 1;  
Matches 44; Conservative 40; Mismatches 62; Indels 55; Gaps 9;

QY 8 ARNTTVKEFILNKTGEVSELKPHRVTVTIQNGKEMSTTVSEEDFILPVYKGELEKGYQ 67  
Db 494 SKNEKSKQ-VLNRNTQLKNEKSKQLPKINVELKEBKQIITKEHN-----ELEKRG-- 542  
QY 68 FDGWEISGPEGKDGAGYVNLKSDTPFKPVPKKIEKKKEBKPTFDVS-----KCKDN 121  
Db 543 -----KGQPKPKVNIELNKEK-VKQVFNESIESKNEKSKQVNVNTQLKNEKSKQL 592  
QY 122 POVN-----HSQLNESHKKE-----DLQRE-----EHSQ-KSOSTKDV 153  
Db 593 PKINVELKEBKSKQISKEHNEBKKEKKIKPKANVELKEBKSKQIIEKHNELKNEKSKQT 652  
QY 154 TATVLDKNNISKSTTNNPNK 174  
Db 653 PKQNVNELKKEKAKHVFNESIK 673

RESULT 29

ADZ79634  
ID ADZ79634 standard; protein; 169 AA.

ADZ79634;

14-JUL-2005 (first entry)

*P. falciparum* merozoite surface protein 3, amino acid residues 212-380.

immune stimulation; fusion protein; merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial; vaccine.

Plasmodium falciparum.

Key Location/Qualifiers

Region 1..169

/note= "Amino acid residues 212-380 of MSP3"

WO2005040206-A1.

06-MAY-2005.

22-OCT-2004; 2004WO-EP012910.

24-OCT-2003; 2003US-00691672.

(INSP ) INST PASTEUR.

Druilbe P;

WPI; 2005-355821/36.

Chimeric molecule useful for preparing vaccine composition against

PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.  
XX  
XX  
PS Claim 2; SEQ ID NO 2; 79pp; English.  
PS  
XX  
XX The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3  
CC protein (amino acid residues 212-380).  
XX  
XX Sequence 169 AA:

```

Query Match      11.5%; Score 103.5; DB 9; Length 169;
Best Local Similarity 25.2%; Pred. NO. 0.13;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7

Qy 41 KEMSSPIVSEEDFILPVYKGBLEKGYQPDGWEISGF--EGKKDAG-----YYINLSKDTFF 93
Db 1 KQASS-----YDYILL : : : : : : : : : : : : : : : : : : : : : :
Qy 94 IKPVFKIKIEKKEEENKPTFDVSKKQNPQVNHSQLN-----ESHKEDLOR 140
Db 42 SKENDVDLDE--KEEAEATESEEELEKNEEETSEISEDEEEEEEKKEEENKKEQEK 100

Qy 141 EHSQKSDSTKQVTVATVLDKNNISKSTTNN 171
Db 101 EGSNNENDOKQWEA-----ONLTSKNNNN 126

```

RESULT 30  
ADW88441  
ID ADW88441 standard: protein: 645 AA.

WP1; 2005-123069/13.

Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
useful for inducing protective immune response in humans against  
*Staphylococcus aureus* infection.

Claim 7; SEQ ID NO 10; 84pp; English.

The present sequence is that of a *Staphylococcus aureus* protein ORF0657n  
hybrid polypeptide denoted 0657nHybrid3. This is an example of claimed  
hybrid polypeptide immunogens of the invention ADW88439-ADW88474 that  
comprise a modified *S. aureus* ORF0657n sequence ADW88433-ADW88438  
containing amino acid substitutions that increase sequence similarity to  
ORF0190 ADW88432. The hybrid polypeptides contain one or more epitopes  
for ORF0657n and ORF0190. They were designed by taking into account the  
similarity and differences between native ORF0657n and ORF0190 protein  
sequences. The invention also provides nucleic acids encoding these  
hybrid polypeptides, and a method for evaluating the ability of an  
immunogen to produce a protective immune response against *Staphylococcus*  
infection using an animal (mouse or rat) model. The hybrid polypeptides  
having therapeutic and diagnostic applications, such as being used to  
provide protective immunity against *S. aureus* infection, being used to  
generate antibodies to detect the presence of *S. aureus*, and being used  
to generate therapeutic antibodies that target *S. aureus*.

Sequence 645 AA;

Query Match	11.4%;	Score 102;	DB 9;	Length 645;	
Best Local Similarity	24.1%;	Pred. No. 1.1;			
Matches	46;	Conservative 33;	Mismatches 68;	Indels 44;	Gaps 8
Qy	1	KIVVKDFA <del>NTVT</del> VKEFILNKTG <del>VE</del> SELKPH <del>RV</del> TTVTIQNGKEMS <del>STIV</del> -----GSEDF 53			
Db	380	KYVMTETND <del>DT</del> WKDFI-----VEGQ <del>RV</del> TVTSKDAKNSRTLLP <del>PI</del> EGTKY <del>YDA</del> 4239			
Qy	54	ILPVYKGELEK <del>YQ</del> YQ <del>FG</del> WEISG <del>FG</del> KDAGV <del>YIN</del> LSKDT <del>FI</del> KPVFKKIEBKGEENK <del>PTF</del> 113			
Db	430	IVKVHV <del>KTID</del> -----YEQY <del>---</del> HVRIIDKDAFTKANT <del>DS</del> KNKKEQ <del>QD</del> NSAKK 474			
Qy	114	DV-----SKKDN <del>PO</del> WN <del>SH</del> QLSN <del>SH</del> RKEDLQ-----REESH <del>SQ</del> SD <del>ST</del> DKVT-ATVLDK <del>NNI</del> 163			
Db	475	EATPAT <del>PS</del> KPT <del>SP</del> VEKESQ <del>QD</del> SK <del>QD</del> KN <del>QD</del> KL <del>PS</del> VEKENDAS <del>ES</del> GK <del>DT</del> PAT <del>PK</del> YK <del>GEV</del> 534			
Qy	164	SKS <del>ST</del> TNN <del>PNK</del> 174			
Db	535	ES <del>ST</del> T--PTK 543			

RESULT 31	
ABB61977	
ID	ABB61977 standard; protein; 564 AA.
XX	
AC	ABB61977;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 12723.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
PR	11-JUL-2000; 2000US-00614150.
XX	

PA (PEXE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX N-PSDB; ABL06080.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 564 AA;  
 SQ  
 Query Match 11.3%; Score 101.5; DB 4; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 1;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EDFILPVYKGELEKGYQPDG-----EISGEGKKDAGYV-----NLSKDTPIK 95  
 Db 78 EDLDTPLSGSRFSK--VFDGWVDEHDEHGDHVDQPSGEALDDHDDHDDHDEDEE 135  
 QY 96 PVFKKIEKKKEENKPT-----FDVSKKQNPQVNHSQLNESHKEDLQREHESKSDS 149  
 Db 136 FUTELELELEEEEFTEDEPADEVEYEDEDEENNA--GENITAEADAEEEDNDND 193  
 QY 150 TKDVTATVLDKNNISSKST 168  
 Db 194 EGTVEATVEATTEATTEAT 212  
 RESULT 32  
 ABR53560  
 ID ABR53560 standard; protein; 1875 AA.  
 XX  
 AC ABR53560;  
 XX  
 DT 20-JUN-2003 (first entry)  
 XX  
 DE Protein sequence #SEQ ID 1985.  
 XX  
 KW Multiprotein complex; eukaryote; drug target; diagnosis.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN EP1258494-A1.  
 XX  
 PD 20-NOV-2002.  
 XX  
 PF 20-DEC-2001; 2001EP-00130253.  
 XX  
 PR 15-MAY-2001; 2001EP-00111774.  
 XX  
 PA (CELL-) CELLZONE AG.  
 XX  
 PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;  
 PI Marzioch M, Schultz JD, Superti-Furga GD;  
 XX WPI; 2003-250078/25.  
 DR N-PSDB; ACC61602.  
 XX

PT New isolated protein complexes useful for diagnosing a disease or  
 PT disorder, or as a target for an active agent of a pharmaceutical,  
 PT preferably a drug target in the treatment or prevention of disease or  
 PT disorder.  
 XX  
 PS Disclosure; SEQ ID NO 1985; 17pp + Sequence Listing; English.  
 XX  
 CC The invention relates to multiprotein complexes from eukaryotes. Proteins  
 CC of the invention and DNA sequences encoding them are given in records  
 CC ABR52568-ABR53903 and ACC610-ACC61944 respectively. The complexes are  
 CC obtainable by using a protein as a bait and isolating the set of proteins  
 CC which is attached thereto from cells. Such protein complexes may comprise  
 CC up to 30 distinct proteins. Protein complexes of the invention are useful  
 CC for diagnosing a disease or disorder, or as a target for an active agent  
 CC of a pharmaceutical, preferably a drug target in the treatment or  
 CC prevention of a disease or disorder. Note: The sequence data for this  
 CC patent is not represented in the printed specification, but is based on  
 CC sequence information supplied by the European Patent Office. The complete  
 CC document is available on CD-ROM  
 XX  
 SQ Sequence 1875 AA;  
 Query Match 11.1%; Score 100; DB 6; Length 1875;  
 Best Local Similarity 24.4%; Pred. No. 7.5;  
 Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;  
 QY 4 VKDFARNTTVKFTILNKDQTEVSELKPHRVTVTTIQNGKMSSTIVSEEDP-----ILPVY 58  
 Db 647 ISQITRESTENMSLANK---BIQDLYDSKSDISIKLGKSSRIILAEERFKLSNTLDT 703  
 QY 59 KGE---LEKGYQPDGWEISGFEKKDA-----GYVINLSKDTPIKPVFKKIEBK--- 104  
 Db 704 KAENDQLKRFYD---LQNTILKQDSKTHETLNEYVSVCKSLKIVETELNLKKEQKL 759  
 QY 105 -----KEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHESKSDSTKDV 153  
 Db 760 VHLKKNLQKELNK-----LSPKQSLRINVTQTLQKREDLLEBTRKSCQKKIDELSDA 815  
 QY 154 TA-----TVLDKNNISSKSTTNNPN 173  
 Db 816 LSELKKTSSQKDHHLKQLEEDNNSN 840  
 RESULT 33  
 ADK64380  
 ID ADK64380 standard; protein; 1875 AA.  
 XX  
 AC ADK64380;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Disease treating protein complex-derived protein #1185.  
 XX  
 KW protein complex; drug target; diagnosis.  
 XX  
 OS Unidentified.  
 XX  
 PN EP1338608-A2.  
 XX  
 PD 27-AUG-2003.  
 XX  
 PF 20-DEC-2002; 2002EP-00102902.  
 XX  
 PR 20-DEC-2001; 2001EP-00130253.  
 XX  
 PA (CELL-) CELLZONE AG.  
 XX  
 PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;  
 PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;  
 PI Michon A, Leutwein C, Rick J;  
 XX WPI; 2003-638460/61.  
 DR N-PSDB; ADK64381.  
 XX







XX SQ Sequence 635 AA;  
Query Match 11.1%; Score 99.5; DB 8; Length 635;  
Best Local Similarity 23.8%; Pred. No. 1.9;  
Matches 44; Conservative 39; Mismatches 67; Indels 35; Gaps 10;  
QY 3 VKKD--FARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEEDFILP--VY 58  
DB 285 IAKOKLFLSLESLKEY-----KGEKVNYEELRF-----NTEPLTSYLENKEKFLVPNIPY 334  
QY 59 KGEL-----EKGYQFD-----GWEISGPEGKGDAGVYNLSKDTPIKPVFKKIEKKKEE 108  
DB 335 KKKILREEDKYSFDEDEBEFGNLLSYNKLKNEVLPVNITTTILKP-----FEQKKIVE 390  
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHS--QKSDSTKDVATVLDKN--NIS 164  
DB 391 D---FNPYSNLDNLEIKKIRLNGSQKQVEOKTSPTPKQETVKEQTKVSGNTQVE 447  
QY 165 SKSTT 169  
DB 448 KKSET 452  
RESULT 37  
ADV8902  
ID ADV89902 standard; protein; 643 AA.  
XX ADV89902;  
XX XX  
DT 24-FEB-2005 (first entry)  
XX Streptococcus agalactiae protein sequence, SEQ ID 2296.  
XX Antibacterial; Vaccine; bacterial infection.  
XX Streptococcus agalactiae.  
XX FR2824074-A1.  
XX 31-OCT-2002.  
XX 26-APR-2001; 2001FR-00005642.  
XX 26-APR-2001; 2001FR-00005642.  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
XX WPI; 2004-101891/11.  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX Claim 6; SEQ ID NO 2296; 2687pp; French.  
XX The present invention relates to novel Streptococcus agalactiae  
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
XX ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
XX agalactiae involved in the synthesis of amino acids, cell membranes,  
XX intermediate (central) metabolism, energetic metabolism, fatty acid and  
XX phospholipid metabolism, nucleotide metabolism including purines,  
XX pyrimidines and/or nucleosides, regulatory functions, replication,  
XX transcription, translation, protein transport, adaptation to atypical  
XX conditions, sensitivity to medicines and/or analogues, functions related  
XX to transposons, biosynthesis of cofactors, prosthetic groups and  
XX transporters, cell membrane proteins and cellular machinery. (I) are  
XX useful for the detection and/or amplification of nucleic acids.  
XX Pharmaceutical composition comprising (I) or (II) are useful for

CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX SQ Sequence 643 AA;  
Query Match 11.1%; Score 99.5; DB 8; Length 643;  
Best Local Similarity 23.8%; Pred. No. 1.9;  
Matches 44; Conservative 39; Mismatches 67; Indels 35; Gaps 10;  
QY 3 VKKD--FARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEEDFILP--VY 58  
DB 293 IAKOKLFLSLESLKEY-----KGEKVNYEELRF-----NTEPLTSYLENKEKFLVPNIPY 342  
QY 59 KGEL-----EKGYQFD-----GWEISGPEGKGDAGVYNLSKDTPIKPVFKKIEKKKEE 108  
DB 343 KKKILREEDKYSFDEDEBEFGNLLSYNKLKNEVLPVNITTTILKP-----FEQKKIVE 398  
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHS--QKSDSTKDVATVLDKN--NIS 164  
DB 399 D---FNPYSNLDNLEIKKIRLNGSQKQVEOKTSPTPKQETVKEQTKVSGNTQVE 455  
QY 165 SKSTT 169  
DB 456 KKSET 460  
RESULT 38  
ADV81155  
ID ADV81155 standard; protein; 643 AA.  
XX AC ADV81155;  
XX XX  
DT 24-FEB-2005 (first entry)  
XX Streptococcus agalactiae protein, SEQ ID 2296.  
XX Streptococcus agalactiae protein, SEQ ID 2296.  
XX Antibacterial; vaccine; bacterial infection.  
XX Streptococcus agalactiae.  
XX WO200292818-A2.  
XX 21-NOV-2002.  
XX 26-APR-2002; 2002WO-IB003059.  
XX 26-APR-2001; 2001FR-00005642.  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX Claim 6; SEQ ID NO 2296; 439pp; French.  
XX The present invention relates to novel Streptococcus agalactiae  
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV83476) and  
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
XX nucleotide sequences encode polypeptides of S. agalactiae involved in the  
XX synthesis of amino acids, cell membranes, intermediate (central)  
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
XX regulatory functions, replication, transcription, translation, protein  
XX transport, adaptation to atypical conditions, sensitivity to medicines

CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
SQ Sequence 643 AA:

```

Query Match      11.1%; Score 99.5; DB 8; Length 643;
Best Local Similarity 23.8%; Pred. No. 1.9;
Matches 44; Conservative 39; Mismatches 67; Indels 35; Gaps 10;

QY      3 VVKD--FARNVTVAKEFLNKDQTVSESLKPHRVTVTIQNGKEMSTIVSEEDFILP--VY 58
      293 IAKDKLSLENSLAEY-----KGEKVYEEIRP-----NTPLTSYLENKKKFLVPNIPY 342

QY      59 KGEL-----RKGYQFD-----GWISGPEGKQAGYVINLSKDTFIPVPPKIKBEKKEE 108
      343 KNKLILREEDKYSPEDDDEERFEGNELLSSYNKLKNEVLPMNITTTILKP---FEQKKIVE 398

QY      109 NKPTFDVSKKQNPQVNHSSQLNESHRKEDLQREHS--QKSDSKDVTATVLDKN--NIS 164
      399 D---FNPYSNLNDLIEIKKIRLNGSQKQVEQKTKSPPTQKETVKBQTEQKVSNGTOEVE 455

QY      165 SKSTT 169
      |||
Db      456 KKSET 460

RESULT 39
ABP55413
ID ABP55413 standard; protein: 1384 AA.

```

04-FEB-2003 (first entry)  
Human MDDT-22 protein SEQ ID NO:22.  
Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
anticonvulsant; antifertility; antiarteriosclerotic; antiasthmatic;  
immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
osteopathic; antihypertic; antiparasitic; antihelminthic; antipsoriatic;  
uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
cancer; developmental disorder; neurological disorder; infection;  
reproductive disorder; autoimmune disorder; inflammatory disorder.

OS	Homo sapiens.
XX	
XX	
PB	WO200278420-A2.
XX	
PD	10-OCT-2002.
XX	
PF	
PP	29-MAR-2002; 2002WO-US009809.
XX	
PR	
PR	30-MAR-2001; 2001US-0280387P.
PR	05-APR-2001; 2001US-0282335P.
PR	13-APR-2001; 2001US-0283663P.
PR	19-APR-2001; 2001US-0285484P.
PR	18-JAN-2002; 2002US-0350702P.
PR	25-JAN-2002; 2002US-0351749P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;
PI	Ramkumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY, Warren BA;
PI	Pignunger DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR, Emerling BM;

	Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AB; WPI; 2003-058385/05. N-PSDB; ABOQ83880.
New	human molecules for disease detection and treatment, useful for diagnosing, treating or preventing autoimmune or inflammatory disorders (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis, cancer or hepatitis.
Claim	1; Page 207-211; 239pp; English.
The present invention describes 23 human molecules for disease detection and treatment (MDDT-1 to 23) (see ABP55392 to ABP55414). The human MDDT-1 to 23 proteins (I) are encoded by the sequences given in ABO83859 to ABO83981. (I) can have various activities depending on the cells and tissues in which they are expressed. These activities include: anti-HIV; antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic; anticonvulsant; antifertility; antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid; cytostatic; hepatotrophic; dermatological; antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoriatic; uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial; virucide; protozoacide; and fungicide. (I) and the polynucleotides encoding them can be used in gene therapy. (I), polynucleotides, agonists and antagonists from the present invention can be used for diagnosing, treating or preventing disorders associated with aberrant expression of MDDT, particularly cell proliferative (e.g. cancer), developmental disorders, neurological disorders, reproductive disorders, or autoimmune/inflammatory disorders, or viral, bacterial, fungal, parasitic, protozoal or helminthic infections. They are also useful in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteins associated with MDDT	
Sequence	1384 AA;
Query Match	11.1%; Score 99.5; DB 6; Length 1384;
Best Local Similarity	21.7%; Pred. No. 5.5;
Matches	45; Conservative 33; Mismatches 67; Indels 63; Gaps 7;
Yq	14 KEPTLNKDTCEVSELKPFRVTVIQNGKMSSITVSEEDPILPVYKELEKGQP----- 68       : :     : :     : :     : :     : :     : :     : :
Dd	1013 KEFIMLQNQEISQLK-KEIFRTQQRKEMBSVMKQEQRYATQYKEADLDGQELALTR 1071     : :     : :     : :     : :     : :     : :     : :
Yq	69 -----DQWEISGFEGKKDA-----GVVINLSK----- 90     : :     : :     : :     : :     : :     : :
Dd	1072 QVNSHTLEAEARHQVOAQREITERLSSELEDNKQLSKERKDANGNHASELGASKVREAH 1131     : :     : :     : :     : :     : :     : :     : :
Yq	91 -----DTFIKVPFKKIEEKKB-----EENKPTFDVSCKKNQPNVNHSLNESHRKE 136     : :     : :     : :     : :     : :     : :     : :
Dd	1132 LEARMQAETIKCLSAEVESLKAYHMEISMISHQENHAWKIS--ADSQSSVVQLNEQLEKA 1189     : :     : :     : :     : :     : :     : :     : :
Yq	137 DLQREHSQKSSTKVDTATVLDKNNI 163     : :     : :     : :     : :     : :     : :
Dd	1190 KLBLEE---AQTTVSNLHOQQVDNRNV 1213     : :     : :     : :     : :     : :     : :
RESULT	40
ABP55393	
ID	ABP55393 standard; protein; 1404 AA.
XX	
AC	ABP55393;
XX	
DT	04-FEB-2003 (first entry)
XX	
DE	Human MDDT-2 protein SEQ ID NO:2.
XX	
Kw	Human; MDDT; molecules for disease detection and treatment; anti-HIV; antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic; anticonvulsant; antifertility; antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid; cytostatic; hepatotrophic; dermatologic; antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoriatic;

KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
KW virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
KW cancer; developmental disorder; neurological disorder; infection;  
KW reproductive disorder; autoimmune disorder; inflammatory disorder.

OS Homo sapiens.

PN WO200278420-A2.

XX 10-OCT-2002.

XX 29-MAR-2002; 2002WO-US009809.

XX 30-MAR-2001; 2001US-0280387P.

PR 05-APR-2001; 2001US-0282335P.

PR 13-APR-2001; 2001US-0283663P.

PR 19-APR-2001; 2001US-0285484P.

PR 18-JAN-2002; 2002US-0350702P.

XX 25-JAN-2002; 2002US-0351749P.

XX (INCY-) INCYTE GENOMICS INC.

PI Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;

PI Ramkumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY, Warren BA;

PI Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR, Emerling BM;

PI Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;

XX WPI; 2003-058385/05.

XX N-PSDB; ABQ83860.

XX New human molecules for disease detection and treatment, useful for

XX diagnosing, treating or preventing autoimmune or inflammatory disorders

XX (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,

XX cancer or hepatitis.

XX Claim 1; Page 172-175; 238pp; English.

XX The present invention describes 23 human molecules for disease detection  
XX and treatment (MDT-1 to 23) (see ABP55392 to ABP55414). The human MDT-1  
XX to 23 proteins (I) are encoded by the sequences given in ABQ83859 to  
XX ABQ83881. (I) can have various activities depending on the cells and  
XX tissues in which they are expressed. These activities include: anti-HIV;  
XX antiallergic; antinflammatory; antianaemic; antiparkinsonian; nootropic;  
XX anticonvulsant; antiinfertility; antiarteriosclerotic; antiasthmatic;  
XX immunosuppressive; antithyroid; cycostatic; hepatotropic; dermatological;  
XX antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
XX osteopathic; antiarthritic; antiparasitic; antihelminthic; antipeoriatic;  
XX uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
XX virucide; protozoacide; and fungicide. (I) and the polynucleotides  
XX encoding them can be used in gene therapy. (I), polynucleotides, agonists  
XX and antagonists from the present invention can be used for diagnosing,  
XX treating or preventing disorders associated with aberrant expression of  
XX MDT, particularly cell proliferative (e.g. cancer), developmental  
XX disorders, neurological disorders, reproductive disorders, or  
XX autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
XX parasitic, protozoal or helminthic infections. They are also useful in  
XX the assessment of the effects of exogenous compounds on the expression of  
XX nucleic acid and amino acid sequences of proteins associated with MDT

XX Sequence 1404 AA;

Query Match 11.1%; Score 99.5; DB 6; Length 1404;

Best Local Similarity 21.7%; Pred. No. 5.7;

Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

QY 14 KEFILNKDVGSEVSELPKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF----- 68

DB 1033 KEFILNQCEISQLK-KEIERTQQRKESVWKQEQYIATQYKEADLDGLRLTRE 1091

QY 69 -----DGEISGPEGKDA-----GYVNLK----- 90

DB 1092 QVQNSHTLAEARHQVQVQRETERLSSELEDMKQL-SEKDAHGHNLAELGASKVREAH 1151

QY 91 -----DTPIKPVFKKIEKKK-----BENKPTFDVSKKDNQPVNHSQLNESHKRE 136  
DB 1152 LEARNQAEIKKLAEVESLKEAYHMEMISHQENHAKWKIS--ADSQKSSVQQLNEQLEKA 1209  
QY 137 DLQREHHSQSDSTKDVVTATVLDKNNI 163  
DB 1210 KLELEE---AQDTVSNLHQVQDRNEV 1233

Search completed: April 24, 2006, 14:50:30

Job time : 111.188 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 27.0537 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTTVKKEFILNK.....ATVLDKNKISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5 COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6 COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/H COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/PCTUS COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/RE COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	773	2	US-09-590-991-8
2	894	99.7	637	2	US-09-107-433-3169
3	894	99.7	2138	2	US-09-583-110-5274
4	615	68.6	117	2	US-08-961-083-68
5	615	68.6	117	2	US-09-536-784-68
6	615	68.6	117	2	US-09-765-271-68
7	615	68.6	117	2	US-09-765-272A-68
8	121	13.5	778	2	US-09-134-001C-3868
9	119	13.3	746	2	US-09-710-279-652
10	104	11.6	347	2	US-09-248-796A-16224
11	97.5	10.9	2468	2	US-09-976-594-726
12	97.5	10.9	2468	2	US-09-538-092-1135
13	97.5	10.9	2522	2	US-09-949-016-10237
14	96	10.7	902	2	US-09-134-001C-3157
15	95	10.6	348	2	US-09-538-092-1316
16	94.5	10.5	277	2	US-09-830-230A-651
17	93.5	10.4	654	2	US-10-172-502-10
18	93	10.4	280	2	US-09-248-796A-17646
19	93	10.4	715	2	US-08-669-286-7
20	93	10.4	715	2	US-09-469-253-7
21	93	10.4	1715	2	US-09-642-146-7
22	92.5	10.3	1702	2	US-08-296-791-5
23	92.5	10.3	1702	2	US-09-839-996-5
24	92.5	10.3	1702	2	US-10-080-505-5
25	92.5	10.3	1702	2	US-10-645-655-5
26	92.5	10.3	1702	4	PCT-US95-10661A-5
27	92	10.3	299	2	US-09-710-279-1888

28	92	10.3	309	2	US-09-134-001C-5667	Sequence 5667, Ap
29	92	10.3	561	1	US-08-360-673-2	Sequence 2, Appli
30	92	10.3	561	5	US-09-640-305-2	Sequence 2, Appli
31	91	10.1	442	2	US-09-134-001C-3033	Sequence 3033, Ap
32	90.5	10.1	743	2	US-08-910-925-3	Sequence 3, Appli
33	90.5	10.1	743	2	US-09-949-016-6261	Sequence 6261, Ap
34	90.5	10.1	758	2	US-09-949-016-8288	Sequence 8288, Ap
35	90	10.0	785	2	US-09-710-279-284	Sequence 264, App
36	89.5	10.0	262	2	US-09-248-796A-21451	Sequence 21451, A
37	89.5	10.0	402	2	US-09-464-483-4	Sequence 4, Appli
38	89.5	10.0	402	2	US-09-414-664-4	Sequence 4, Appli
39	89.5	10.0	511	2	US-09-198-452A-509	Sequence 509, App
40	89.5	10.0	511	2	US-09-438-185A-475	Sequence 475, App
41	89.5	10.0	529	2	US-09-464-483-2	Sequence 2, Appli
42	89.5	10.0	553	2	US-09-248-796A-16588	Sequence 16588, A
43	89.5	10.0	299	2	US-09-328-352-6567	Sequence 6567, Ap
44	89	9.9	299	2	US-09-134-001C-3148	Sequence 3148, Ap
45	88	9.8	443	2	US-09-134-001C-3148	Sequence 3148, Ap

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; TYPE: PRT  
; LENGTH: 773  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score 897;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 8.9e-83;		
Matches 174;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG	60	
Db	600	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG	659	
QY	61	ELEKGYQDGEISGFEKCKDAGVYINLSKDTFTKPVFKLEEKKKEENKPTFDVSKKD	120	
Db	660	ELEKGYQDGEISGFEKCKDAGVYINLSKDTFTKPVFKLEEKKKEENKPTFDVSKKD	719	
QY	121	NPQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNKISSKSTNNPNK	174	
Db	720	NPQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNKISSKSTNNPNK	773	

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD/ROM 1809660
/ COMPUTER: <Unknown>
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/ 085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 3169:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 637 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...637
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
/
/ US-09-107-433-3169
/
/
/ Query Match 99.7%; Score 894; DB 2; Length 637;
/ Best Local Similarity 99.4%; Pred. No. 1.4e-82;
/ Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 60
/ DB 430 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 489
/
/ QY 61 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKKD 120
/ DB 490 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKKD 549
/
/ QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 174
/ DB 550 NPQVNSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 603
/
/ RESULT 3
/ US-09-583-110-5274
/ Sequence 5274, Application US/09583110
/ Patent No. 6699703
/
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al.
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
/ FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
/ FILE REFERENCE: PATH00-07A
/ CURRENT APPLICATION NUMBER: US/09/583,110
/ CURRENT FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/107,433
/ PRIOR FILING DATE: 1998-06-30
/ PRIOR APPLICATION NUMBER: US 60/085,131
/ PRIOR FILING DATE: 1998-05-12
/ PRIOR APPLICATION NUMBER: US 60/051,553
/
```

```
/ PRIOR FILING DATE: 1997-07-02
/ NUMBER OF SEQ ID NOS: 5322
/ SEQ ID NO 5274
/ LENGTH: 2138
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ US-09-583-110-5274
/
/ Query Match 99.7%; Score 894; DB 2; Length 2138;
/ Best Local Similarity 99.4%; Pred. No. 7.3e-82;
/ Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 60
/ DB 1931 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 1990
/
/ QY 61 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKKD 120
/ DB 1991 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKKD 2050
/
/ QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 174
/ DB 2051 NPQVNSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 2104
/
/ RESULT 4
/ US-08-961-083-68
/ Sequence 68, Application US/08961083
/ Patent No. 6159469
/
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340F2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-961-083-68
/
/ Query Match 68.6%; Score 615; DB 2; Length 117;
/ Best Local Similarity 100.0%; Pred. No. 3.7e-55;
/ Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 58 YKGELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSK 117
/
```

Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60  
QY 118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 174  
Db 61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/536,784  
; FILING DATE: 30-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-536-784-68

Query Match 68.6%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 3.7e-55;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 117  
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60  
QY 118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 174  
Db 61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,271  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/536,784  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-271-68

Query Match 68.6%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 3.7e-55;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 117  
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60  
QY 118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 174  
Db 61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 7

US-09-765-272A-68  
; Sequence 68, Application US/09765272A  
; Patent No. 6929930  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and  
; NUMBER OF SEQUENCES: 454  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: Dell Latitude C610  
; OPERATING SYSTEM: Windows 2000  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:



```
; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: FB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match      68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3,7e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YGELKSGYQFDGWEISGEGKKDAGYVNLSDTFIKVPFKIEBKGEENKPTPDVSK 117
DB 1 YGELKSGYQFDGWEISGEGKKDAGYVNLSDTFIKVPFKIEBKGEENKPTPDVSK 60

QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPK 174
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPK 117

RESULT 8
US-09-134-001C-3868
; Sequence 3868, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3868
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3868

Query Match      13.5%; Score 121; DB 2; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.0012;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;

QY 6 DPAR--NTTVKEFILNKDTEGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYK---- 59
DB 601 DFKSVPNVQDREYQKAEDSNAQSLKP-----ITIGNGKQIKQOSVKSQKVLPHSKVMLM 656
QY 60 --GELEKGYQFDGW---EISGFE-----GKKDAGYVIN--LSKDTFIKVPFKIEBKTEKK 105
DB 657 TDGELTWP-DMTGWTEDVLAFEDLTKIKVSTKNGFVTNQISKQGIK----- 705
QY 106 EENKPTFDVS-----KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNN 161
DB 706 ---NKDKIEVLSAEDTDDQEKTDDESDSKSKDKADEHDSNTSSSTKN-----DKS 756
```

```
QY 162 NISSKSTNN 171
DB 757 NADSKNDSDD 766
```

## RESULT 9

```
US-09-710-279-652
; Sequence 652, Application US/09710279
; Patent No. 6703432
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 652
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-652
```

```
Query Match      13.3%; Score 119; DB 2; Length 746;
Best Local Similarity 27.0%; Pred. No. 0.0018;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 20 KDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYK-----GELEKGYQFDGW-- 71
DB 585 EDSVNAQSLKP----ITIGNGKQIKQOSVKSQKVLPHSKVMLMTDGLTWP-DMTGWTK 639

QY 72 -EISGFE-----GKKDAGYVIN--LSKDTFIKVPFKIEBKGEENKPTFDVS----K 117
DB 640 EDVLAFEDLTKIKVSTKNGFVTNQISKQGIK-----NKDKIEVLSAED 686

QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 171
DB 687 TDDQEKTDDESDSKSKDKADEHDSNTSSSTKN-----DKSNADSKNDSDD 734
```

## RESULT 10

```
US-09-248-796A-16224
; Sequence 16224, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224
```

```
Query Match      11.6%; Score 104; DB 2; Length 347;
Best Local Similarity 19.9%; Pred. No. 0.022;
Matches 39; Conservative 42; Mismatches 83; Indels 32; Gaps 6;

QY 4 VKDFARN-TTVKEFILNKDTEGEVSELKPHR-----VTVTIQNGKEMSSITVSE 50
```

```

; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Db 151 LKTELKSPFLKQKFLSYDQVRKLEMTREKLVOAWKDCGLDFDVKQLKSLPLMIDE 210
QY 51 EDFILPVPKGELEKGYQPDGWSISGF-----EGKKDAGYVINLSKDTFIKPVFKKIBEK 104
Db 211 IDTIPRIYND--KQWVVVATSSLQNVQDLSSESSEIGWEDDLSENYRTGPFVKTLDQL 268
QY 105 KEENKPTPDVSKKDNPNVNSQLN-----ESHRKEDLQREHSQKSDSTKVTVATVL 158
Db 269 REE-----WKASKEQANPKKEBENLNQKPVAKQKQKPNSTKKQKQTKQKQTKKITKPKT 323
QY 159 DKNNISSKSTNNPNK 174
Db 324 SKRMLEGISTNLIINK 339

RESULT 11
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Db 14 KEFILNKDGTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVVKKDKPVKTKETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
QY 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBKKEENKPTFDVSKKKDNPNVNSQLNESH 133
Db 624 QATDVKPKAAKKTETKTKVKP-----EDKKEEKKPKKEVAKKEDTKPI---KKEBKP 675
QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
```

```

; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Db 14 KEFILNKDGTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVVKKDKPVKTKETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
QY 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBKKEENKPTFDVSKKKDNPNVNSQLNESH 133
Db 624 QATDVKPKAAKKTETKTKVKP-----EDKKEEKKPKKEVAKKEDTKPI---KKEBKP 675
QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 13
US-09-949-016-10237
; Sequence 10237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10237
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10237

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2522;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Db 14 KEFILNKDGTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 638 EKVVKKDKPVKTKETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 677
QY 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBKKEENKPTFDVSKKKDNPNVNSQLNESH 133
Db 678 QATDVKPKAAKKTETKTKVKP-----EDKKEEKKPKKEVAKKEDTKPI---KKEBKP 729
QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 730 KKEEVKKEVKKEIKKEBKEKPKKEV 754

RESULT 14
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
```





```
/ ATTORNEY/AGENT INFORMATION:
/ NAME: MURPHY Jr., Gerald M.
/ REGISTRATION NUMBER: 28,977
/ REFERENCE/DOCKET NUMBER: 230-110P
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 205-8000
/ TELEFAX: (703) 205-8050
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 715 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-469-253-7

Query Match 10.4%; Score 93; DB 2; Length 715;
Best Local Similarity 21.0%; Pred. No. 0.77;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;

QY 3 VVKDPARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGEL 62
Db 239 IADISNRKMAKLYMVDASGSM-----RVTVAEENPFPSMALLSECFILD--HGAA 290
QY 63 EKGYPDGEWEISGFEGK-----KDAGYVILSKDTFIK-----PVFKK----- 100
Db 291 KQIFVWKGKDANPQERKAAMKTAEBFLQOMYSKNTQIQVLPEGGETPIFKQFFKDRDK 350
QY 101 -----IEKKKEENKPTFVSKKKNPOV--NHSQLNESHKEDLOREHS 144
Db 351 DQSDGFGKVVYVTEKVAQIKQIPFDA3KLHSSPQMAAQHNMYDDGSGKVEIWRVENN 406

RESULT 21
US-09-642-146-7
/ Sequence 7, Application US/09642146
/ Patent No. 6271353
/ GENERAL INFORMATION:
/ APPLICANT: NAKAMURA, SEIJI
/ APPLICANT: SAKURAI, TAKASHI
/ TITLE OF INVENTION: GENE ENCODING ADSEVERIN
/ NUMBER OF SEQUENCES: 18
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Birch, Stewart, Kolaach & Birch, LLP
/ STREET: P.O. Box 747
/ CITY: Falls Church
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22040-0747
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/642,146
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/669,286
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: MURPHY Jr., Gerald M.
/ REGISTRATION NUMBER: 28,977
/ REFERENCE/DOCKET NUMBER: 230-110P
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 205-8000
/ TELEFAX: (703) 205-8050
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 715 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
```

```
/ MOLECULE TYPE: protein
/ US-09-642-146-7

Query Match 10.4%; Score 93; DB 2; Length 715;
Best Local Similarity 21.0%; Pred. No. 0.77;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;

QY 3 VVKDPARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGEL 62
Db 239 IADISNRKMAKLYMVDASGSM-----RVTVAEENPFPSMALLSECFILD--HGAA 290
QY 63 EKGYPDGEWEISGFEGK-----KDAGYVILSKDTFIK-----PVFKK----- 100
Db 291 KQIFVWKGKDANPQERKAAMKTAEBFLQOMYSKNTQIQVLPEGGETPIFKQFFKDRDK 350
QY 101 -----IEKKKEENKPTFVSKKKNPOV--NHSQLNESHKEDLOREHS 144
Db 351 DQSDGFGKVVYVTEKVAQIKQIPFDA3KLHSSPQMAAQHNMYDDGSGKVEIWRVENN 406

RESULT 22
US-08-296-791-5
/ Sequence 5, Application US/08296791
/ Patent No. 6245337
/ GENERAL INFORMATION:
/ APPLICANT: St. Geme III, Joseph W.
/ APPLICANT: Falkow, Stanley
/ TITLE OF INVENTION: Haemophilus Adherence and Penetration
/ TITLE OF INVENTION: Protein
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert
/ STREET: 4 Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: United States
/ ZIP: 94111-4187
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/296,791
/ FILING DATE: 25-AUG-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trecartin, Richard F.
/ REGISTRATION NUMBER: 31,801
/ REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1702 amino acids
/ TYPE: amino acid
/ TOPOLOGY: unknown
/ US-08-296-791-5

Query Match 10.3%; Score 92.5; DB 2; Length 1702;
Best Local Similarity 23.9%; Pred. No. 2.9;
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDPARNT---TVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSE---EDFI 54
Db 1206 VVSKNQNTENTTDOPTEREKTAKEVTEKQE--PPQVSAQSPKQSQSETVQPVAVLESN 1263
QY 55 LPVYKGELEKGYQPDGWEISGFEGKDA-GVVINLSKDTFIKPVFKLEKKKEENKPTFF 113
Db 1264 VPTVNAEEVQALQQTQTSATVSTVSTQAPENSINTGSATAITETAESKDKPKQTAASTE 1323
```

QY 114 DVSKKONQVNHSHKEDLQREHS---QKSDTKDVTATVLDKNNISSKSTTN 170  
Db 1324 DASQHKANTVADNSVANNSSSEPKSRRRRSISQPOETSABETTAASDTETIADNSKRS 1383

QY 171 NPNK 174  
Db 1384 KPNR 1387

RESULT 23  
US-09-839-996-5  
; Sequence 5, Application US/09839996  
; Patent No. 6642371  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; PROTEIN  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839,996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELE: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-839-996-5

Query Match 10.3%; Score 92.5; DB 2; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDFARNT---TVKGFILNKDTGEVSELKPHRTVTIIONGKEMSSIVSE---EDFI 54  
Db 1206 VVSKNTENTTDOPTEREKTAKEVTEKTE---PPQVASQASPKQEQSETVQPOAVLESEN 1263

QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKIIEKKKEENKPTF 113  
Db 1264 VPTVNNAEVQAQLQTSATVSTKQAPENSINTGSATATETAEKSKPQTETAASTE 1323

QY 114 DVSKKONQVNHSHKEDLQREHS---QKSDTKDVTATVLDKNNISSKSTTN 170  
Db 1324 DASQHKANTVADNSVANNSSSEPKSRRRRSISQPOETSABETTAASDTETIADNSKRS 1383

QY 171 NPNK 174  
Db 1384 KPNR 1387

RESULT 25  
US-10-645-655-5  
; Sequence 5, Application US/10645655  
; Patent No. 6815182  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; PROTEIN  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/645,655  
; FILING DATE: 20-Aug-2003  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791

;  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
;  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-10-645-655-5  
  
Query Match 10.3%; Score 92.5; DB 2; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;  
  
QY 2 IVVKDFARNT-----TVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSIVSE---EDFI 54  
Db 1206 VVSKNQNTENTDQPTEREKTAKVETEKQ--PPQVASQASPKQEQSETVQPQAVLESEN 1263  
  
QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKIEEKEENKPTF 113  
Db 1264 VPTVNNAEVQAQLQTQTSATVSTKQAPENSINTGSATATETAEKSDKPKQTETAASTE 1323  
  
QY 114 DVSKKNDPQVNHSQLNESHKEDLQREHS---OKSDSTKQVTVATVLDKNNISSKSTTN 170  
Db 1324 DASQKANTVADNSVANNSESSEPKSRRRRSISQPKETSAEETTAASTDETTIADNSKES 1383  
  
QY 171 NPNK 174  
Db 1384 KPNR 1387  
  
RESULT 26  
PCT-US95-10661A-5  
; Sequence 5, Application PC/TUS9510661A  
; GENERAL INFORMATION:  
; APPLICANT: Washington University, et al.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Teet, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10661A  
; FILING DATE: 16-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: FP-59941/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299

;  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; PCT-US95-10661A-5  
  
Query Match 10.3%; Score 92.5; DB 4; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;  
  
QY 2 IVVKDFARNT-----TVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSIVSE---EDFI 54  
Db 1206 VVSKNQNTENTDQPTEREKTAKVETEKQ--PPQVASQASPKQEQSETVQPQAVLESEN 1263  
  
QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKIEEKEENKPTF 113  
Db 1264 VPTVNNAEVQAQLQTQTSATVSTKQAPENSINTGSATATETAEKSDKPKQTETAASTE 1323  
  
QY 114 DVSKKNDPQVNHSQLNESHKEDLQREHS---OKSDSTKQVTVATVLDKNNISSKSTTN 170  
Db 1324 DASQKANTVADNSVANNSESSEPKSRRRRSISQPKETSAEETTAASTDETTIADNSKES 1383  
  
QY 171 NPNK 174  
Db 1384 KPNR 1387  
  
RESULT 27  
US-09-710-279-1888  
; Sequence 1888, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-1888  
  
Query Match 10.3%; Score 92; DB 2; Length 299;  
Best Local Similarity 24.4%; Pred. No. 0.29;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
  
QY 19 NKDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFI---LPVYKGELEKGYQFDGWEI 73  
Db 33 NKDT-EKSDKXTHRIISLIPSNTEILYRLGIGEDIVGVSTVDYDPKDVKKKKQFDAMWL 91  
  
QY 74 SGFE-----GKDGAGYVINLSKDTFIKPV-----FKKIEEKK 105  
Db 92 NKEELIKAKPOLILAHESQKNSAGIKVLSKDKGVVYVVKDAQSIDETDTFKSIGOUT 151  
  
QY 106 EENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSQ 145  
Db 152 DREKQAKELVDETKHNVEKIINSVPKHHKKQEVFMEVSSK 191  
  
RESULT 28  
US-09-134-001C-5667  
; Sequence 5667, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:



```

;      TOPOLGY: linear
;      MOLECULE TYPE: protein
US-08-360-673-2

Query Match          10.3%; Score 92; DB 1; Length 561;
Best Local Similarity 23.5%; Pred. No. 0.7;
Matches 43; Conservative 32; Mismatches 72; Indels 36; Gaps 8;

QY      4 VKDFARNTTVKSFILNKDTGEVSELSKPHRVTTTIOGKEMSS-----TIVSEEDFIL 55
Db      90 VNDIANKIENRVIIFKKDASADEVKFHQLVSVEHAKALGLADHDHPFTATSGEHSF 149

QY      56 PYYKGELEKYGFODGWISGGFGKKDAGVINLSKDT--FIK--PVFKIIEBK----- 105
Db      150 GKVAHSLEGGIQ-DSPDIAG---SLSGVGVPYTKKEVIDFIRSPIVEFEEDSMVFNS 204

QY      106 -BEENKPTFDVSKGDKNPOVNHSQLNESHKRKDLQREHSQKSDSTKVTTATVLDKN-NI 163
Db      205 FNTQNSAPGLARIASHREKLNLGSFN-----XYLYDDAGKGVTAVVDITGVNV 253

QY      164 SSK 166
Db      254 NHK 256

```

Patent No. RE37447  
GENERAL INFORMATION:  
APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
PREPARATION AND USE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Releasee #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/640,305  
FILING DATE: 16-Aug-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,673  
FILING DATE: 06-FEB-1995  
APPLICATION NUMBER: WO PCT/FR93/00623  
FILING DATE: 23-JUN-1993  
APPLICATION NUMBER: FR 92/07785  
FILING DATE: 25-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST92040-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 561 amino acids

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-640-305-2

Query Match      10.3%; Score 92; DB 5; Length 561;
Best Local Similarity 23.5%; Pred. No. 0.7;
Matches 43; Conservative 32; Mismatches 36; Indels 8; Gaps 8;

QY 4 VKDFARTTTVEKILNKDTCGEVSELKPHRVTVTIQNGKEMSS-----TIVSEEDFIL 55
Db 90 VNDIANKIPNRYIIVFKDASADEVKFQHELVSVEHAKALGSLADHPPTATSGHSEF 149
QY 56 PYYKGELEGYQPDGWEISGPEGKDGAGVNLKDT--PIK--PVFKKIEKK----- 105
Db 150 GVKAHSLGGIQ--DSFDIAG----SLSGYGVGYFTKEVIDFIRSRPLVFEVFEEDSMVFNS 204
QY 106 -EENKPTPDVSKKDNPNVHSQLNESHKEDLQREHSQKSDSKDVTATVLDKN-NI 163
Db 205 FNTQNSAPGLARISHREKLNLSFN-----KYLYDDDGAGVTAIVDTGVNV 253
QY 164 SSX 166
Db 254 NHK 256

RESULT 31
US-09-134-001C-3033
; Sequence 3033, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3033
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3033

Query Match      10.1%; Score 91; DB 2; Length 442;
Best Local Similarity 21.7%; Pred. No. 0.64;
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;

QY 20 KDTGEVSELKPHRVTVT--IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64
Db 90 EDT-EINPVAQPEVWTQIEKGDFEATVVEPVKLDGYKGLIEKQETDLSDEELQ 148
QY 65 -----GYQFDGWEISG--FEGKDGAGVNLKSDTFIK 95
Db 149 ESIDHSLSLHAEVMVKEDGAVNGDVTNIDFSG-SVDGEFDFGQAGEYDLEIGSGSFIP 207
QY 96 PVFKKIEEKEENK-----PTFDVS-----KKKNQPNVHSQLNE- 131
Db 208 GFEEQIEGKMTGDEKDVVTFPEYHAEELAGKEATFKTKVNEIKPKQVPELNDEITANEL 267
QY 132 -----SHRKEDLQREHSQKSDSKDVTATVLDKNNTSSKSTTN 170
Db 268 DSDAENVDEKENLRKELSEQKATEAENT-----EKEEAINKATEN 308

RESULT 32
US-08-910-925-3
; Sequence 3, Application US/08910925
```

```
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SP-LICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0365 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1684847
; US-08-910-925-3

Query Match      10.1%; Score 90.5; DB 2; Length 743;
Best Local Similarity 23.8%; Pred. No. 1.5;
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 27 ELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGS-----LEKGYQFDGWEI--SQF 76
Db 281 EARPR-----QSMKEKHQVVRNEEHKAEQEGKVAQREBELVETGQHNNDVEIEEAGE 335
QY 77 EGKDGAGVNLKSDTFIKPVFKKIEEKKBEENKPTFDVSKKDNPNVHSQLNESHKRB 136
Db 336 EEEKEIGIVHSDAE-----KEQEEEEKQKQEMEVKMEEE-----TEVRESEKQ 378
QY 137 DLQREHS---QKSDTKDVTAT--TVLDKNNTSSKSTTN 171
Db 379 DSQPEEVMVLEVMENVKHVIADQEVMETNVRVESVEPSEN 418

RESULT 33
US-09-949-016-6261
; Sequence 6261, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6261  
; LENGTH: 743  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6261

Query Match 10.1%; Score 90.5; DB 2; Length 743;  
Best Local Similarity 23.8%; Pred. No. 1.5;  
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 27 ELKPHRVTTIQNGKEMSTIYSEEDFILPVYKGE-----LEKGYQPDGWEI--SGF 76  
Db 281 EAPRPR-----QSMKEKHQVVRNEEHKABQEGKVAQREBELVETGNOHNVIEEAGE 335  
QY 77 EGKKGAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKDNPOVNHSQLNESHKXE 136  
Db 336 EEEKEIGIVHSDAE-----KEQEEBKQKEKVMWEE-----TEVRSEKQQ 378  
QY 137 DLQREHS---QKSDSTKDVT--TVLDKNNISSKSTNN 171  
Db 379 DSQPEVMDVLEMVNVKHVIADQEVMTNRVESVEPSN 418

## RESULT 34

US-09-949-016-8288  
; Sequence 8288, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8288  
; LENGTH: 758  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-8288

Query Match 10.1%; Score 90.5; DB 2; Length 758;  
Best Local Similarity 28.1%; Pred. No. 1.5;  
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 23 GEVSELKPHRVTTIQNGKEMSTIYSE--EDFILPVYKGELEKGYQF-----DGWE--- 72  
Db 399 GBELQQLSGSQL-----HGKSDSPNVYTEKKEKTAILERLUTELERLKTPEQRSDLWRLY 453  
QY 73 -----ISGPEGKKDAG-----YVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKCD 120  
Db 454 VRKQDNGKQGTGDKKGGGRSHRAKNKSKETPLGVS-----KETFDAMKNST 501  
QY 121 NPOVNH-----SQLNESHKEDLQREHSQKSD--STKDVATVLDK 160  
Db 502 KEFVRHHEKIKQAQEA--VKENLKPFSDSVKSTFRHFKDTTKNIIDE 547

## RESULT 35

US-09-710-279-264  
; Sequence 264, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 264  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-264

Query Match 10.0%; Score 90; DB 2; Length 785;  
Best Local Similarity 23.7%; Pred. No. 1.8;  
Matches 42; Conservative 33; Mismatches 78; Indels 24; Gaps 8;  
QY 7 PARNTTVKBFILNKDQTVSELSKPHRVTTIQNGKEMS-STIVSEEDFILPVYKGELEKG 65  
Db 615 YKRTFVSEYRLNREGGKIK----TATITERGNHIVITTVTGEDLMVVTNAGVI--- 667

QY 66 YQFDGWEISGPEGKKDAGY-VINLSKDTFIKPVFKKIEKKKEEN-----KPTPDV 115  
Db 668 IRLDVHDISQ-NGRAAQGVRLMKLGQGFVSTAKVNEEDDNEENADEAQQSTTTTADV 726  
QY 116 SKKKNPQVNHSQLNESHKEDLQREH-HSQKSDSTKDVTATVLDKNNISSKSTNN 171  
Db 727 EEWVD-----DQTPGNAIHTEGDAEMESVEFPENDRIDRIDQFMDRVNEDIESASDN 779

## RESULT 36

US-09-248-796A-21451  
; Sequence 21451, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 21451  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-21451

Query Match 10.0%; Score 89.5; DB 2; Length 262;  
Best Local Similarity 26.1%; Pred. No. 0.44;  
Matches 43; Conservative 26; Mismatches 43; Indels 53; Gaps 9;

QY 20 KDTGVSEL-KPHRVTVT-----IQNGKEMSTIYSEEDF----- 53  
Db 52 KSTPKTSPLRKPFPKPTVTVKMAKSRPPSVTNTPEIKPKBSSESPISESDFLEMD 111  
QY 54 -----ILPVYKGELEKGYQFDGWEISGPEGKKDAGVVLNLSKDTFIKPVFKKIEKKKEEN 109  
Db 112 KLTEVEVPIIE-----FNFDNYD--NDEDKKEE--VVKNSKNE--NQNITKGMELKPKPKP 160

QY 110 KPTFDVSKKDNPNVNHSQLNESHKEDLQREHSQKSDSTKDTVT 154  
Db 161 KPEETSKTKVRPQ-----PRK---QKKQKPLSEETVDTLT 193

## RESULT 37

US-09-464-483-4  
; Sequence 4, Application US/09464483  
; Patent No. 6228617  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth J.  
; TITLE OF INVENTION: No. 6228617el tig  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/464,483  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,339  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10085  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:

INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 402 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-464-483-4

Query Match 10.0%; Score 89.5; DB 2; Length 402;  
Best Local Similarity 22.1%; Pred. No. 0.79;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 21 DTGEVSELKPHRYVT-IONGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64  
Db 53 DETDIKPAQPEVSVTQIEKGDFIFEAIVTVEPVKLGDKYGLKLEIKQETELSDDELQ 112  
QY 65 -----GYQFDGWEISG--FEGKDGAGVYVNLKSDTFIKP 96  
Db 113 AIDHSLGLAEMVVKEDGVVNGDVTNIDFSG-SVDGEFEGGQAGYDLEIGSGSFI-P 170  
QY 97 VFK-----KIEEKE-----EE--NKPTFDVS-----KKKNPNVNHSQLNE- 131  
Db 171 GFEEQLGKMKVDEKDVVTPPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230  
QY 132 -----SHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTTN 170  
Db 231 DAEANTVDEYKLNRLKLAEQKATDAENV-----EKEEAITKATDN 271

## RESULT 38

US-09-414-664-4  
; Sequence 4, Application US/09414664

; Patent No. 6242249  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth J.  
; TITLE OF INVENTION: No. 6242249el tig  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/414,664  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/999,339  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10085  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:

INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 402 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-414-664-4

Query Match 10.0%; Score 89.5; DB 2; Length 402;  
Best Local Similarity 22.1%; Pred. No. 0.79;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 21 DTGEVSELKPHRYVT-IONGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64  
Db 53 DETDIKPAQPEVSVTQIEKGDFIFEAIVTVEPVKLGDKYGLKLEIKQETELSDDELQ 112  
QY 65 -----GYQFDGWEISG--FEGKDGAGVYVNLKSDTFIKP 96  
Db 113 AIDHSLGLAEMVVKEDGVVNGDVTNIDFSG-SVDGEFEGGQAGYDLEIGSGSFI-P 170  
QY 97 VFK-----KIEEKE-----EE--NKPTFDVS-----KKKNPNVNHSQLNE- 131  
Db 171 GFEEQLGKMKVDEKDVVTPPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230  
QY 132 -----SHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTTN 170  
Db 231 DAEANTVDEYKLNRLKLAEQKATDAENV-----EKEEAITKATDN 271

## RESULT 39

US-09-198-452A-509  
; Sequence 509, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 509  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-509

Query Match 10.0%; Score 89.5; DB 2; Length 511;  
Best Local Similarity 24.5%; Pred. No. 1.1;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 94 IKPVFKKIEKKKEENKPTD-----VSKKKNPQVNHSQLNESHRK 135  
Db 95 VKGVFKKTPQARPEVSSPLPSHVQHGQRLPGLGFRDRIQKSENPEADLGKQKRSYSD 154  
Qy 136 EDLQREHSQSDSTKDVTATVLDKNNISSKSTT 169  
Db 155 GDLDRVGHDSNEDSTEDSR---EGGEPSSKSSS 185

RESULT 40

US-09-438-185A-475  
; Sequence 475, Application US/09438185A  
; Patent No. 6822071  
; GENERAL INFORMATION:  
; APPLICANT: Stephens, Richard  
; APPLICANT: Mitchell, Wayne  
; APPLICANT: Kalman, Sue  
; APPLICANT: Davis, Ronald  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence  
; FILE REFERENCE: 018941-000411US  
; CURRENT APPLICATION NUMBER: US/09/438,185A  
; PRIOR FILING DATE: 2002-03-13  
; PRIOR APPLICATION NUMBER: US 60/108,279  
; PRIOR FILING DATE: 1998-11-12  
; PRIOR APPLICATION NUMBER: US 60/128,606  
; PRIOR FILING DATE: 1999-04-08  
; NUMBER OF SEQ ID NOS: 1074  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 475  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
; FEATURE:  
; OTHER INFORMATION: CPn0473  
US-09-438-185A-475

Query Match 10.0%; Score 89.5; DB 2; Length 511;  
Best Local Similarity 24.5%; Pred. No. 1.1;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 94 IKPVFKKIEKKKEENKPTD-----VSKKKNPQVNHSQLNESHRK 135  
Db 95 VKGVFKKTPQARPEVSSPLPSHVQHGQRLPGLGFRDRIQKSENPEADLGKQKRSYSD 154  
Qy 136 EDLQREHSQSDSTKDVTATVLDKNNISSKSTT 169  
Db 155 GDLDRVGHDSNEDSTEDSR---EGGEPSSKSSS 185

Search completed: April 24, 2006, 15:03:41  
Job time : 29.0537 secs

**THIS PAGE BLANK (uspio)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 87.1946 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTTVKKEFILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	773	4	US-10-067-385-8
2	897	100.0	2119	3	US-09-769-744A-28
3	897	100.0	2140	4	US-10-282-122A-73670
4	897	100.0	2140	5	US-10-472-928-1180
5	894	99.7	637	5	US-10-617-320-3169
6	615	68.6	117	3	US-09-765-272-68
7	615	68.6	117	6	US-11-106-649-68
8	121	13.5	778	4	US-10-724-972A-5663
9	111.5	12.4	707	4	US-10-282-122A-52942
10	111.5	12.4	775	4	US-10-282-122A-70721
11	110.5	12.3	647	5	US-10-691-672A-3
12	109	12.2	665	3	US-09-820-843A-107
13	108.5	12.1	973	5	US-10-732-923-18783
14	108	12.0	188	5	US-10-691-672A-7
15	107.5	12.0	470	5	US-10-739-930-6262
16	107.5	12.0	948	5	US-10-732-923-4286
17	105	11.7	3127	5	US-10-732-923-22588
18	104.5	11.6	903	4	US-10-282-122A-52328
19	104	11.6	540	5	US-10-732-923-22820
20	103.5	11.5	169	5	US-10-691-672A-2
21	102	11.4	1419	5	US-10-732-923-4285
22	101.5	11.3	564	6	US-11-097-143-12723
23	101.5	11.3	1434	4	US-10-732-923-4235
24	101	11.3	933	4	US-10-437-963-166606
25	100.5	11.2	1350	5	US-10-732-923-8760
26	100.5	11.2	1373	5	US-10-732-923-16976
27	100	11.1	1875	4	US-10-369-493-22285

28	100	11.1	1875	5	US-10-732-923-3334	Sequence 3334, Ap
29	100	11.1	1875	5	US-10-732-923-3335	Sequence 3335, Ap
30	99.5	11.1	1384	4	US-10-473-576-22	Sequence 22, Appl
31	99.5	11.1	1404	4	US-10-473-576-2	Sequence 2, Appli
32	99.5	11.1	3124	5	US-10-732-923-22709	Sequence 22709, A
33	98.5	11.0	2060	4	US-10-381-598A-2	Sequence 2, Appli
34	98	10.9	441	5	US-10-510-812-14	Sequence 14, Appl
35	98	10.9	891	4	US-10-437-963-166609	Sequence 166609,
36	98	10.9	1419	5	US-10-732-923-4288	Sequence 4288, Ap
37	97.5	10.9	1184	4	US-10-282-122A-53254	Sequence 53254, A
38	97.5	10.9	2468	4	US-10-755-889-615	Sequence 615, App
39	97.5	10.9	2468	5	US-10-489-740-216	Sequence 216, App
40	97.5	10.9	2519	5	US-10-450-763-46995	Sequence 46995, A
41	96.5	10.8	898	4	US-10-425-115-205148	Sequence 205148,
42	96	10.7	893	4	US-10-282-122A-70831	Sequence 70831, A
43	96	10.7	902	4	US-10-724-972A-6781	Sequence 6781, Ap
44	96	10.7	1939	5	US-10-732-923-3340	Sequence 3340, Ap
45	94.5	10.5	234	3	US-09-764-870-367	Sequence 367, App

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Chol, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 897;	DB 4;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 2.7e-67;	Mismatches 0;	Indels 0; Gaps 0;
Matches 174;	Conservative	0;		
QY	1	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKG	60	
DB	600	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKG	659	
QY	61	ELEKGYQDGMWISFGEGKODAGYVINSKOTFIKVPKCIIEKKEENKPTFDVSKKD	120	
DB	660	ELEKGYQDGMWISFGEGKODAGYVINSKOTFIKVPKCIIEKKEENKPTFDVSKKD	719	
QY	121	NPQVNHSQLNESHREKDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK	174	
DB	720	NPQVNHSQLNESHREKDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK	773	

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Phillip M  
; TITLE OF INVENTION: Proteins



```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-769-744A-28

Query Match      100.0%; Score 897; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9.5e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60
Db 1912 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1971

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120
Db 1972 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 2031

QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2032 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-282-122A-73670

Query Match      100.0%; Score 897; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60
Db 1933 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120
Db 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 2052

QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2053 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.8+01)
; US-10-472-928-1180

Query Match      100.0%; Score 897; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60
Db 1933 KIVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120
Db 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 2052

QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2053 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
;
; US-10-617-320-3169
;
;
; Query Match 99.7%; Score 894; DB 5; Length 637;
; Best Local Similarity 99.4%; Pred. No. 3.9e-67; Mismatches 0; Indels 0; Gaps 0;
; Matches 173; Conservative 1;
;
; QY 1 KIVKDFARNTTVKGFILNKDTGEVSELKPHRTVTIQQKEMSTIVSEEDFILPVYKG 60
; DB 430 KIVKDFARNTTVKGFILNKDTGEVSELKPHRTVTIQQKEMSTIVSEEDFILPVYKG 489
;
; QY 61 ELEKGYQFDGWEISGFEKGKADAGYVNLKDTFIKPVFKIEEKEEENKPTFDVSKKD 120
; DB 490 ELEKGYQFDGWEISGFEKGKADAGYVNLKDTFIKPVFKIEEKEEENKPTFDVSKKD 549
;
; QY 121 NFQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 174
; DB 550 NFQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 603
;
; RESULT 6
; US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
;
; US-09-765-272-68
;
; Query Match 68.6%; Score 615; DB 3; Length 117;
; Best Local Similarity 100.0%; Pred. No. 2e-44;
; Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 58 YKGELEKGYQFDGWEISGFEKGKADAGYVNLKDTFIKPVFKIEEKEEENKPTFDVSK 117
; DB 1 YKGELEKGYQFDGWEISGFEKGKADAGYVNLKDTFIKPVFKIEEKEEENKPTFDVSK 60
;
; QY 118 KQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 174
; DB 61 KQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 117
;
; RESULT 7
; US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 68
```

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      68.6%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-44;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YGELKGYQFGWEISGEGKDDAGYVNLKDTPIKVPFKKIEKKKEENKPTFDVSK 117
Db 1 YGELKGYQFGWEISGEGKDDAGYVNLKDTPIKVPFKKIEKKKEENKPTFDVSK 60
;
QY 118 KDNPNVNSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNISKSTTNNPNK 174
Db 61 KDNPNVNSQLNESHKEDLQREHHSQKSDSKDVTATVLDKNISKSTTNNPNK 117
;
RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S. epidermidis
US-10-724-972A-5663

Query Match      13.5%; Score 121; DB 4; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.15;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;

QY 6 DPAR--NTTVKGFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVVK--- 59
Db 601 DFKVPNVQGVDEVQKAEKDSVNAQSLKP-----ITIGNGKQIKQOQSVKSGTKVLPKPHSKVLM 656
;
QY 60 --GELEKGYQFGW-----EISGPE-----GKKDAGYVIN--LSKDTFTKVPFKKIEKK 105
Db 657 TDGELTWP--DMTGTKEDVLAFEDLTAKIVSTKNGFVTVNQISKGIIK----- 705
;
QY 106 EENKPTFDVS-----KKDNPNVNSQLNESHKEDLQREHHSQKSDSKDVTATVLDKN 161
Db 706 ---NKKIKIEVLSAEDTDDQKEDSDSNKSKKQKADEHSDNTSSSTKN-----DKS 756
;
QY 162 NISSKSTTN 171
Db 757 NADSKNDSDD 766
;
RESULT 9
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
```

FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (359)..(359)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (385)..(385)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (388)..(388)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (396)..(396)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (400)..(400)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (402)..(402)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (404)..(404)  
OTHER INFORMATION: X=any amino acid  
US-10-282-122A-52942

Query Match 12.4%; Score 111.5; DB 4; Length 707;  
Best Local Similarity 26.6%; Pred. No. 0.83;  
Matches 47; Conservative 28; Mismatches 63; Indels 39; Gaps 7;  
QY 2 IVKQDARNTTKKFI--LNDKQGVSELPKPHRVTTIQNGKMSSTIVSEEDFILPVYK 59  
Db 484 ISIEDAER-GVKEEDSNQDIDGVVEDK-----TTDKYEDS---NKSDIIEPNK 532  
QY 60 GLEKGYQDGFWEISGFEGKQAGYVINLSKDTFKPVFKKIEEKKKEEN--KPTFDVSK 117  
Db 533 KSKKAKLFG-----FIKKQNEVEQEENLNDISPDILDK 569  
QY 118 KQNPQVNSQLNESHKEDLQREHSQKSDSTQVTVTLQKNNISSKSTNNPNK 174  
Db 570 PVENNQVKEIEQNLKE-IKQEPHQHIEERSVKIEKPINNILDEKVVSSNNESK 625

RESULT 10  
US-10-282-122A-70721  
Sequence 70721, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 70721  
LENGTH: 775  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70721

Query Match 12.4%; Score 111.5; DB 4; Length 775;  
Best Local Similarity 24.9%; Pred. No. 0.93;  
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;  
QY 20 KDTGEVSELPKPHRVTTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGM-- 71  
Db 614 EDSVNAQLKP-----ITIGNGKQIKQSVKSTKVLPHSKVLMMDTDELTMP-DMTGWTYK 668  
QY 72 -BISGFE-----GKKQAGYVIN--LSKDTFKPVFK-----KIEKKKEENKPTF 113  
Db 669 EDVLAPEDLTIKVSTKNGFVTVNSISKQIKNKQKLEVSLSAEDTDDQKTDDESS 728  
QY 114 DVSKKKQNPQVNSQLNESHKEDLQREHSQKSDSTQVTVTLQKNN 162  
Db 729 DKSKKQKQVDEDSNASSSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 11  
US-10-691-672A-3  
Sequence 3, Application US/10691672A  
Publication No. US20050112133A1  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
FILE REFERENCE: 02356.0085  
CURRENT APPLICATION NUMBER: US/10/691,672A  
CURRENT FILING DATE: 2003-10-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 3.3  
SEQ ID NO 3  
LENGTH: 647  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Peptide  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)..(647)  
OTHER INFORMATION: GLURP MSP3 fusion protein  
US-10-691-672A-3

Query Match 12.3%; Score 110.5; DB 5; Length 647;  
Best Local Similarity 22.8%; Pred. No. 0.91;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 14 KEFILNKDTGVSELPKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELE 63

```
Db 418 EBAVSEKNAHETVE---HEETVQESNPEKADNGNVSONNENLNEFV-----ESE 468
Qy 64 K-----GQFDF-GWEISGF--EGKQDAG-----YVINLSKDTFKPVFKKIE 102
Db 469 KSEHARSKAKASSYDILGMEFGGVPEHKKEENMLSHLYVSSKKENISKENDVDLD 528
Qy 103 EKKEENKPTFDVSKKKNPQVNSQLN-----ESHKREDLQREHSQKSDS 149
Db 529 E-KEEAETEEELKEKEETESEISEDEBEKEEKEEKEEKEEKEEKEEKEEENDDQ 587
Qy 150 TKDVTATVLDKNNISSKSTNN 171
Db 588 KKOMEA-----QNLISKNNNN 604

RESULT 12
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match 12.2%; Score 109; DB 3; Length 665;
Best Local Similarity 26.1%; Pred. No. 1.3;
Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

Qy 5 KDFARNTTVKERTLNKDTGEVSELKPHRTV-VTIQNGKEMSSITVSEEDFILPVYKGELE 63
Db 114 KDNNNNGTKLEEKNNKNSDL--HQNELMLQSGK-----NEQDI-----NKN 158
Qy 64 KGQFDFGWEI--SGFEGKQDAGYVINLSKDTFKPVFKKIEKKE-----EENKP 111
Db 159 KGKQ-----DISNSAENKQD-----VKEGVKLEEKKEEKISDDHKVEENCK 202
Qy 112 TPD-----VSKKKNPQVNSQLNESHKRDQOR-EHSQKSDSTKDVATVLDKNNISSK 166
Db 203 SDDHKVEENKSDDHKVEENKSDDHKIEVKKVEEHEDEE-----DKKEKSE 253
Qy 167 STTNPNK 174
Db 254 NKNKDNK 261

RESULT 13
US-10-732-923-18783
; Sequence 18783, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 18783
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(973)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-18783

Query Match 12.1%; Score 108.5; DB 5; Length 973;
Best Local Similarity 21.7%; Pred. No. 2.2;
Matches 38; Conservative 44; Mismatches 80; Indels 13; Gaps 5;

Qy 5 KDFARNTTVKEF-ILNKDTGE-----VSELKPHRTVVTIQNGKEMSSITVSEEDFILPV 57
Db 568 KDIRDNKSSSFTLVNNNTNNGNNSYDQKLETYINNNNDQGGECNSTYKT-----LIQH 623
Qy 58 YKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFKPVFKKIEKKEENKPTFDVSK 117
Db 624 RSGSKESGKTHHTKINNISRERHARGYVSLSKONNIEVRNDNIEVRNDNIEV 683
Qy 118 KKNPQVNSQLNESHKREDL-QREHSQKSDSTKDVATVLDKNNISSKSTNN 171
Db 684 RNDNIEMRNDSINDQNEKNISFNSFKSEYT-FVTATSNKSKKDDNINKSSND 737

RESULT 14
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match 12.0%; Score 108; DB 5; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.32;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

Qy 37 IQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 89
Db 15 VLKAEASS-----YDYIL-----GHEFGGVPEHKKEENMLSHLYVSSKD 55
Qy 90 KDTFKPVFKKIEKKEE-----ENKPTFDVSKKKNPQVNSQLNESHKRE 136
Db 56 KENISKENDVDLDEKEEAEETEELKEKNEETESEISEDEBEKEEKEEKEEENDKKK 115
Qy 137 DLQREHSQKSDSTKDVATVLDKNNISSKSTNN 171
Db 116 EQEKEQSNENNDQKKOMEA-----QNLISKNNNN 145

RESULT 15
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```



```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52328
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52328

Query Match          11.6%; Score 104.5; DB 4; Length 903;
Best Local Similarity 21.9%; Pred. No. 4.4;
Matches 44; Conservative 40; Mismatches 62; Indels 55; Gaps 9;

QY 8 ARNTTVKPFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQ 67
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 494 SKNEKSKQ-VLNRNTQLKNEKSKQLPKINVELKEEKNKQIIKEHN-----ELEKG-- 542

QY 68 PDGWEISGEGKQDAGYVNLKDTPIKVPFKIEKKEEKNKPTFDVS-----KKQDN 121
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 543 -----GKQPKQDKVNIELNKEK-VKQVFNEISIESKNEKSKQVNVNRTQLKNEKSKQL 592

QY 122 PQVN-----HSQNESHRKE-----DLORE-----EHSQ-KSDSTKDV 153
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 593 PKINVELKEEKSQKISKEHNELRKEKKIKPRANVELKEEKSQKIIKEHNEKNEKSKQT 652

QY 154 TATVLDKNNISSKSTTNNPNK 174
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 653 PKVNVLEKKEKAGHVFNESEIK 673

RESULT 19
US-10-732-923-22820
; Sequence 22820, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22820
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-22820

Query Match          11.6%; Score 104; DB 5; Length 540;
Best Local Similarity 25.3%; Pred. No. 2.6;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TTVKPFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFG 70
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 27 TTVKA-VVEETKVEEDSKP-----EGVEKSASFKESDFFADLKESEKK----- 70

QY 71 WEISGEGKQDAGYVNLKDTPIKVPFK--IIEKKEEKNKPTFDVSKKQDNQVNHIS 127
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 71 -ALSDLKSLKEAIVDN-----TLTKTKCKESSPMKEGSEVVKPEAEVKKCKE--BAABE 123

QY 128 QLNESHKEDLOREHSQKSDSTKQVTVTLVDKNNISS 165
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
```

```
Db 124 KVEBEKKSEAVVTEAPKAETVEAVVTEIIPKEEVT 161

RESULT 20
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691.672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match          11.5%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.67;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 41 KEMSTIVSEEDFILPVYKGELEKGFQDGWEIISGF--EGKKDAG-----YVINLSKDTF 93
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1 KEASS-----YDIYL-----GWFGGVPPEHKKENMLSHLYVSSKDKENI 41

QY 94 IKPVFKITEEKEEENKPTFDVSKKQNPQVNHSQLN-----ESHKEDLQR 140
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 42 SKENDDVLDL-KEBEAEETELEEKNEEETESEISEDEEESEEEKEEKEKQEQEK 100

QY 141 EEHSQKSDSTKQVTVTLVDKNNISSKSTTN 171
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 101 EQSNENDQKKDMEA-----QNLISKNQNNN 126

RESULT 21
US-10-732-923-4285
; Sequence 4285, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4285
; LENGTH: 1419
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-4285

Query Match          11.4%; Score 102; DB 5; Length 1419;
Best Local Similarity 22.0%; Pred. No. 13;
Matches 52; Conservative 37; Mismatches 77; Indels 70; Gaps 11;

QY 4 VKDFARNTTVKPFILNKDTGEVSEL-----KPHRVTVTIQNGKMSSTIVSEEDFI 54
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1019 VFDMSNMGFFNYIINLGGYPEELEYNKAGDKIHYCTLSLQIKDKIOKFLNEETPL 1078

QY 55 LPVY-----KGELEKGYQFDGWEISGEGKKD-----AGYVINLSKDTFIKPVFKK 100
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1079 KTKGYGSFEKISLAINNSIDHY-----FSHMKDNLRVICPGRYNVAASSTLAVKIIGKR 1134
```





```
QY 169 TNN 171
Db 380 SDS 382

RESULT 25
US-10-732-923-8760
; Sequence 8760, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8760
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(1350)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8760

Query Match 11.2%; Score 100.5; DB 5; Length 1350;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 46; Conservative 37; Mismatches 74; Indels 43; Gaps 10;

QY 5 KDPARNTTVKBFILNKDTGEVSELKPHRVTVTQNGK-----EWSSTIVSDEDFILPVY 58
Db 846 KKLKNSSTIKKFLKNN-----KKHMFIDLGERKSKWKVMNTACTKTNKKKAILYGW 897

QY 59 KGELEKGYQFDGW-----EISGFEGKDGAGYVNLSDTKFVKPKIEKKKEE---N 109
Db 898 KAARGGHDPQFPNNEKLEKEIEKKWN-NYHINQOK---IKBIIESQTEKEDFEKIIN 953

QY 110 KPTFDVSKKKD-----NPOV-NHSQLNESHKEDLORE-----EHSQKSDSTKQVT 154
Db 954 IHEFLTHAKNIYLNILNPNLINDNSIGDSIEEKDVKSDIDEDGNEGKKSGEFNGI- 1012

QY 155 ATVLDKNNISSKSTNNPNK 174
Db 1013 VNIYSKSNVSGSSETRNDSK 1032

RESULT 26
US-10-732-923-16976
; Sequence 16976, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 16976
; LENGTH: 1373
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-16976

Query Match 11.2%; Score 100.5; DB 5; Length 1373;
Best Local Similarity 21.7%; Pred. No. 16;
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 16 FILNKDTGEVSE---LKPHRVTVTQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGW 71
Db 418 VYNEDRGVWEESIIMLPHPHVQILJ-----LSATV-----PNY-----LEFADW 457

QY 72 -----EISGFEGKK-----DAGYVNLSDTKFVKPKIEKKKEE 108
Db 458 VGFTKQKEVISISTKCRPVPLLLHYTVYDSVYLVMDKKNKFFSSAFKBIYVYKIREKQEAN 517

QY 109 NKPIFDVSKKDNQVNHSQLNESH-RKEDLOREHSHQSDSTKQVDTATVLDKNNISSKS 167
Db 518 NKNTKQITSGSNTSSNLKNNYDYSKRYLTTTNNKENDNT-----QNNINNNN 568

QY 168 TNN 171
Db 569 NNNN 572

RESULT 27
US-10-369-493-22285
; Sequence 22285, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22285
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match 11.1%; Score 100; DB 4; Length 1875;
Best Local Similarity 24.4%; Pred. No. 26;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

QY 4 VKDFARNTTVKBFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEEDF-----ILPVY 58
Db 647 ISQITRESTENMSLINK---BIQDIYDSKSDISIKLGKESRILAEERFKLLSNTLOIT 703

QY 59 KGE---LEKGYQFDGWSEISGFEGKKDA-----GYVNLSDTKFVKPKIEK--- 104
Db 704 KAENDQLRRFDY-----LQNTILKCDSTHETLNEIVSCKSLSIVETELMLKBEQKUR 759

QY 105 -----KEEENKPTFDVSKKDNQVNHSQLNESH-RKEDL---QREHSHQSDSTKQV 153
Db 760 VHLEKVLKQELNK-----LSPEKDSLRIMVTQLOTQKEREDLLEETRSCKCKIDLEDA 815

QY 154 TA-----TVLDKNNISSKSTNNPN 173
Db 816 LSELKKETSQKDHIIKQLEEDNNSN 840

RESULT 28
US-10-732-923-3334
; Sequence 3334, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
```

```

; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3334
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3334

Query Match      11.1%; Score 100; DB 5; Length 1875;
Best Local Similarity 24.4%; Pred.No.26;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

Qy 4 VKDPARVTVKEPFLNKDTGEVSELKPHRVTVTTQNGKMSSTIVSEDP-----ILPVY 58
Db 647 ISQITRESTENMSLNK---EIQLYDSKSDISIKLGKSSRILABERFKLLSNTLDLT 703
Qy 59 KGE---LEKGYOPDGEISGPEGKDA-----GYVINLSKDTPIKVPFKKIEK--- 104
Db 704 KAENDQLRKRPDY-----LQNTILKQDSKTHETLNLEYVSCSKSIVSETLLNLKBEQKLR 759
Qy 105 -----KBEENKPTPVDVSKKKNQPNVHSQLNESHR-KEDL---OREHSQKSDSTKDV 153
Db 760 VHLEKNLKQELNK---LSPKDSLRIMVTQLTQLKEREDLLEETRSCKQKIDLEDA 815
Qy 154 TA-----TVLDKNNISSKSTTNPN 173
Db 816 LSELKGETSQDHHIKQLEEDNNSN 840

RESULT 29
US-10-732-923-3335
; Sequence 3335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3335
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3335

Query Match      11.1%; Score 100; DB 5; Length 1875;
Best Local Similarity 24.4%; Pred.No.26;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

Qy 4 VKDPARVTVKEPFLNKDTGEVSELKPHRVTVTTQNGKMSSTIVSEDP-----ILPVY 58
Db 647 ISQITRESTENMSLNK---EIQLYDSKSDISIKLGKSSRILABERFKLLSNTLDLT 703
Qy 59 KGE---LEKGYOPDGEISGPEGKDA-----GYVINLSKDTPIKVPFKKIEK--- 104
Db 704 KAENDQLRKRPDY-----LQNTILKQDSKTHETLNLEYVSCSKSIVSETLLNLKBEQKLR 759
Qy 105 -----KBEENKPTPVDVSKKKNQPNVHSQLNESHR-KEDL---OREHSQKSDSTKDV 153
Db 760 VHLEKNLKQELNK---LSPKDSLRIMVTQLTQLKEREDLLEETRSCKQKIDLEDA 815
Qy 154 TA-----TVLDKNNISSKSTTNPN 173
Db 816 LSELKGETSQDHHIKQLEEDNNSN 840

RESULT 30
US-10-473-576-22
; Sequence 22, Application US/10473576
; Publication No. US20040101884A1

```

Db 1190 KLELEE---AQTIVSNLHQVQDRNEV 1213

RESULT 31

US-10-473-576-2

; Sequence 2, Application US/10473576

; Publication No. US20040101884A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE CORPORATION

; APPLICANT: LU, DYUNG AINA M.

; APPLICANT: ARVIZU, CHANDRA S.

; APPLICANT: GANDHI, AMENA R.

; APPLICANT: HAFALIA, APRIL J.A.

; APPLICANT: DING, LI

; APPLICANT: LU, YAN

; APPLICANT: RAMKUMAR, JAYALAXMI

; APPLICANT: SWARNAKAR, ANITA

; APPLICANT: TANG, Y. TOM

; APPLICANT: YUE, HENRY

; APPLICANT: TRAN, BAO

; APPLICANT: LEE, SOO YUEN

; APPLICANT: WARREN, BRIDGET A.

; APPLICANT: NGUYEN, DANIEL B.

; APPLICANT: THANGAVELU, KAVITHA

; APPLICANT: YAO, MONIQUE G.

; APPLICANT: ELLIOTT, VICKI S.

; APPLICANT: BAUGHN, MARIAH R.

; APPLICANT: EMERLING, BROOKE M.

; APPLICANT: LAL, PREETI G.

; APPLICANT: GIETZEN, KIMBERLY J.

; APPLICANT: BECHA, SHANYA D.

; APPLICANT: MAROUI, JOSEPH P.

; APPLICANT: KABLE, AMY E.

; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT

; FILE REFERENCE: PF-0921 USN

; CURRENT APPLICATION NUMBER: US/10/473,576

; CURRENT FILING DATE: 2003-09-29

; PRIOR APPLICATION NUMBER: PCT/US02/09809

; PRIOR FILING DATE: 2002-03-29

; PRIOR APPLICATION NUMBER: US 60/280,387

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: US 60/282,335

; PRIOR FILING DATE: 2001-04-05

; PRIOR APPLICATION NUMBER: US 60/286,663

; PRIOR FILING DATE: 2001-04-13

; PRIOR APPLICATION NUMBER: US 60/285,484

; PRIOR FILING DATE: 2001-04-19

; PRIOR APPLICATION NUMBER: US 60/350,702

; PRIOR FILING DATE: 2002-01-18

; PRIOR APPLICATION NUMBER: US 60/351,749

; PRIOR FILING DATE: 2002-01-25

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PERL Program

; SEQ ID NO 2

; LENGTH: 1404

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No: 3125036CD1

US-10-473-576-2

Query Match 11.1%; Score 99.5; DB 4; Length 1404;

Best Local Similarity 21.7%; Pred. No. 20;

Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

Qy 14 KEPLINKDTGCEVSELKPHRYVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF-----68

Db 1033 KEFIMLQNEQIEISQLK-KEIERTQQRKMESEVYKQEQIATQYKEADLQGLRLTRE 1091

Qy 69 -----DGWEISGFEGKKDA-----GVYINLSK-----90

Db 1092 QVQNSHTELAEARHQVQAQREIURLSSELEDMKQLSKEDKANGHNLABELGASKVREAH 1151

Qy 91 -----DTFIKPVFKKIEBKKE-----EENKPTFDVSKKDNPOVNHSQLNESHKRE 136

Db 1152 LEARMQAEIKVLSAEVESLKEAYHMEISHQENHAKWKIS--ADSQSSVQQLNEQLEKA 1209

Qy 137 DLQREHSQKSDSTKQVATVTLVDIQNI 163

Db 1210 KLELEE---AQTIVSNLHQVQDRNEV 1233

RESULT 32

US-10-732-923-22709

; Sequence 22709, Application US/10732923

; Publication No. US20050108791A1

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

; FILE REFERENCE: 38-15(52796)C

; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10

; PRIOR APPLICATION NUMBER: 10/310,154

; PRIOR FILING DATE: 2002-12-04

; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 22709

; LENGTH: 3124

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-10-732-923-22709

Query Match 11.1%; Score 99.5; DB 5; Length 3124;

Best Local Similarity 23.1%; Pred. No. 54;

Matches 53; Conservative 35; Mismatches 76; Indels 65; Gaps 9;

Qy 9 RNTTVK--EPLINKDTGEVSELKP-----HRTVTITQNG-----KEMSSSTIVSE 50

Db 1950 RNTYKKSDDLNDKNGESKFKGNTSYVLESPLHLIGDIVDNNIKRKKKKKIKTIIVSD 2009

Qy 51 EDFILPVYKGELEKGYQFDGWEISG---FEKKDAGYVINLSKDTFIKPVFKKIEBKKEE 107

Db 2010 DMFTSPVNIKEYNYNEQERKEIVGNLSYDKTKICPFKFTKSGRIKK--NKIEKKKK 2067

Qy 108 ENKPTF-----DVSKKKDNPOVNHSQL-----129

Db 2068 EYNNFLYNDYSSVPKYGNENNPFVIVIRERKDFKDFHFNPFKFLHYNPMK 2127

Qy 130 NESHKRE---DLQREHSQKSDSTKD-VTATVLDKNNISSKSTTNNPK 174

Db 2128 NKNKNKQNVNRRNEYPNYTSSTSKDGVSYNPLSDLSFSSDNEYSSDNE 2176

RESULT 33

US-10-381-596A-2

; Sequence 2, Application US/10381596A

; Publication No. US20040014178A1

; GENERAL INFORMATION:

; APPLICANT: Biostapro AB

; TITLE OF INVENTION: von Willebrand factor-binding proteins from

; TITLE OF INVENTION: Staphylococci

; FILE REFERENCE: 110059600

; CURRENT APPLICATION NUMBER: US/10/381,596A

; CURRENT FILING DATE: 2003-07-02

; PRIOR APPLICATION NUMBER: SE 0003573-3

; PRIOR FILING DATE: 2000-10-04

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2060

; TYPE: PRT

; ORGANISM: Staphylococcus lugdunensis

US-10-381-596A-2

Query Match 11.0%; Score 98.5; DB 4; Length 2060;

Best Local Similarity 25.4%; Pred. No. 39;

```
Matches 45; Conservative 26; Mismatches 67; Indels 39; Gaps 9;
QY 1 KIVVKDPARNTTVKFEFLNKDTGEVSELKPHRVTVTI---QNGKEMSTIVSEEDFILPV 57
Db 1887 KVTED-KITTVYVERIQSVNDAPRVDIDELKTIIVYDTNGRE-----IVPS 1934
QY 58 YKGELEKGYQFDG--WEISFEGKKDAGYVNLNLSKDTFKVPVKIE-----KKEBE 108
Db 1935 RKQLPP-BQFQGDQYTGHK-----TEKDGITTYIYKVENAVPAKQLKKTGH 1983
QY 109 NKPTFDVSKKDNPNVHNSQLNESHKE--DLQREHSHQSDSKDVTATVLDKNNI 163
Db 1984 N--TQSESQKHTPQVKQLVXKYNVKEQBSIEKSEHTDMHVSLEPGETGTANKGL 2038

RESULT 34
US-10-510-812-14
; Sequence 14, Application US/10510812
; Publication No. US20050176097A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Michael Dolberg
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Olsen, Peter Bjarke
; TITLE OF INVENTION: Improved Bacillus host cell
; FILE REFERENCE: 10296.204-US
; CURRENT APPLICATION NUMBER: US/10/510,812
; CURRENT FILING DATE: 2004-10-07
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 14
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-812-14

Query Match 10.9%; Score 98; DB 5; Length 441;
Best Local Similarity 31.6%; Pred. No. 6.5;
Matches 43; Conservative 18; Mismatches 35; Indels 40; Gaps 9;
QY 6 DPA-RNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEK 64
Db 302 DVAFENPTWKBFY-----AKGAQVKGNE--TVTVDKGKTEVGIVTSEKAFSMPVYKGE-BK 354
QY 65 GYQFDGWEISFEGKKDAGYVNLNLSKDTFKVPVKIE-----KKEENKPTF---D 114
Db 355 NYKAK-----VTLNKDELTA PV-KKGEKVGILTASYKGEKDYGLGSD 397

QY 115 VS-----KKKDNPNV 125
Db 398 VSGVNLVTKEDDEKAN 413

RESULT 35
US-10-437-963-166609
; Sequence 166609, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 166609
```

```
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_65300C.1.pap
US-10-437-963-166609

Query Match 10.9%; Score 98; DB 4; Length 891;
Best Local Similarity 22.8%; Pred. No. 15;
Matches 42; Conservative 39; Mismatches 75; Indels 28; Gaps 7;
QY 3 VKDPAFNTTVKFEFLNK-----DTGVSSELKPHRVTVTIQNGKEMSTIVSEED 52
Db 199 LVSNFSCRDLVEFEFVLLKIRPLSSGWDITLGIPEDCPSLPPFV-----VSSDLVIPED 253
QY 53 FILPVK--GELEKGYQFDGWEISFEGKKDAGYVNLNLSKDTFK---PVPKKIREKKEE 107
Db 254 MVRPVKFLGPTYTKDEHLKFLTQ--NGKRQVRVAGSDIPVRVHPEIVPALKKKKSV 311
QY 108 ENKPTFDVSKKDNPNVHNSQLNESHREDLQREHSHQSDSKDVTATVLDKNNISSKS 167
Db 312 --KPSSD---DDDDVLEEDVDEIGEENAEESGEGADEENDSSDDSDSSNSSDN 365
QY 168 TTNN 171
Db 366 SSDS 369

RESULT 36
US-10-732-923-4288
; Sequence 4288, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4288
; LENGTH: 1419
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-4288

Query Match 10.9%; Score 98; DB 5; Length 1419;
Best Local Similarity 20.0%; Pred. No. 27;
Matches 50; Conservative 34; Mismatches 68; Indels 98; Gaps 10;
QY 4 VKDPAFNTTVKFEFLNKDTGEVSEL-----KPHRVTVTIQNGKEMSTIVSEEDFI 54
Db 1019 VFDSSNMGNFNYIILNGLGGYPPELEYDNKXKDKHYCTLSLQBIKKDIQKFLNEETFL 1078
QY 55 LPVYKGELEKGYQFDGWEISFEGKKDAGYVNLNLSKDTFKI-----94
Db 1079 KTKY-----GYSEKISLAINNSIDHYFSHMKONLVICEPGSNMVA 1122
QY 95 -----KPVFKKI-----EKKGEENKPTFDVSKKDNPNV 125
Db 1123 SSTLAVKIIGKRPTFGIIMLKDLKHYDPLNPAQENKKQDETNIEN--NDNNNDNN 1181
QY 126 HSQLESHEKE-----DL-----QREHSHQSDSKDVTATVLDK--NNIS 164
Db 1182 DNNINNNNNKGGGQGNIMNDLIITSTNDSTSKNDHS--SSQVIQNVSCITRDKEGDNIK 1240
QY 165 -SKSTNNPN 173
Db 1241 INTHTINPN 1250

RESULT 37
```

```

; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 615
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-615

Query Match 10.9%; Score 97.5; DB 4; Length 2468;
Best Local Similarity 24.8%; Pred. No. 60;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6

QY 14 KEFLNKDTCGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGELEKGYQFDGWEI 73
DB 584 EKVMVKDKPKVKTETKESVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 74 SGFEGKKDAGYVNLNKDTPFKPVFKIEKKKEENKPTFDVSKKKONPQVNHSQLNESH 133
DB 624 QATDVKPKAAKKTVKKETVKP-----EDKKEEKPKKEVAKKEDKTPI---KKEKP 675

QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
DB 676 KKEEVKKEVKKEIKKEEKKEPKKEV 700

RESULT 39
US-10-489-740-216
; Sequence 216, Application US/10489740
; Publication No. US20050112574A1
; GENERAL INFORMATION:
; APPLICANT: Bionomics Limited
; TITLE OF INVENTION: P9
; FILE REFERENCE: Angiogenesis PCT
; CURRENT APPLICATION NUMBER: US/10/489,740
; CURRENT FILING DATE: 2004-03-15
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 216
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-489-740-216

Query Match 10.9%; Score 97.5; DB 5; Length 2468;
Best Local Similarity 24.8%; Pred. No. 60;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6

QY 14 KEFLNKDTCGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGELEKGYQFDGWEI 73
DB 584 EKVMVKDKPKVKTETKESVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 74 SGFEGKKDAGYVNLNKDTPFKPVFKIEKKKEENKPTFDVSKKKONPQVNHSQLNESH 133
DB 624 QATDVKPKAAKKTVKKETVKP-----EDKKEEKPKKEVAKKEDKTPI---KKEKP 675

QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
DB 676 KKEEVKKEVKKEIKKEEKKEPKKEV 700

RESULT 40
US-10-450-763-46995
; Sequence 46995, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:

```

```
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 46995
; LENGTH: 2519
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1040)..(1091)
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1919)..(2122)
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by PFam,
; OTHER INFORMATION: accession name MAP1B_neuraxin, E-value=1.9e-59, PFam score of 190
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2519)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-46995
```

```
Query Match          10.9%; Score 97.5; DB 5; Length 2519;
Best Local Similarity 24.8%; Pred.No.61;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 14 KEFILNKDTGVSSELKPHRVTVYTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 635 EKVMVKDKPKVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 674

QY 74 SGFEGKKDAGYVINLSKDTFIPKVPFKIEKKKEENKPTFDVSKKQDNPQVNHSQLNESH 133
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 675 QATDVKPKAAKEKTVKGETKVKP-----EDKKKEKEKPKKEVAKCEDKTPI---KKBEKP 726

QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 727 KKEEVKKEVKKEIKKEEKKPKKEV 751
```

Search completed: April 24, 2006, 15:43:38  
Job time : 89.1946 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 13.6242 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KLVKDFARNTTVKEFILNK.....ATVLDKNNISKSTNNPNK 174

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA New.\*  
1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /SIDSS/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /SIDSS/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	13.3	746	US-10-793-626-652	Sequence 652, App
2	112.5	12.5	354	US-11-189-817-2	Sequence 2, Appli
3	110.5	12.3	651	US-11-128-660-1	Sequence 1, Appli
4	95	10.6	886	US-11-087-099-11456	Sequence 11456, A
5	93.5	10.4	501	US-10-485-517-381	Sequence 381, App
6	93.5	10.4	645	US-10-485-517-244	Sequence 244, App
7	93	10.4	493	US-11-096-568A-3070	Sequence 3070, App
8	93	10.4	493	US-11-096-568A-3071	Sequence 3071, App
9	93	10.4	510	US-11-096-568A-3069	Sequence 3069, App
10	92	10.3	299	US-10-793-626-1888	Sequence 1888, App
11	90	10.0	785	US-10-793-626-264	Sequence 264, App
12	89.5	10.0	1036	US-11-096-568A-28315	Sequence 28315, A
13	89.5	10.0	1070	US-11-096-568A-28314	Sequence 28314, A
14	89.5	10.0	1276	US-11-096-568A-28313	Sequence 28313, A
15	89.5	10.0	1694	US-11-052-554A-83	Sequence 83, Appl
16	87	9.7	140	US-11-096-568A-4771	Sequence 4771, App
17	86	9.6	443	US-11-188-298-1015	Sequence 1015, App
18	86	9.6	700	US-11-188-475-74	Sequence 74, Appl
19	85.5	9.5	734	US-11-072-512-2272	Sequence 2272, App
20	85	9.5	394	US-10-793-626-552	Sequence 552, App
21	85	9.5	472	US-10-793-626-658	Sequence 658, App
22	85	9.5	636	US-10-485-517-170	Sequence 170, App
23	85	9.5	700	US-11-196-475-66	Sequence 66, Appl
24	85	9.5	708	US-11-196-475-76	Sequence 76, Appl
25	85	9.5	720	US-10-793-626-2058	Sequence 2058, App

#### ALIGNMENTS

##### RESULT 1

US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	13.3%	Score 119;	DB 6;	Length 746;
Best Local Similarity	27.0%	Pred. NO. 0.044;		
Matches	47;	Conservative	24;	Mismatches 57;
				Indels 46;
				Gaps 9;
QY	20	KDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYK-----GSELEKGYQFDGW---	71	
DB	585	EDSVNAQSLKP-----ITTNGKQKQKQSVKSGTKVLPHSKVLMYDGLTWP-DMTGWTK	639	
QY	72	-EISGFE-----GKKDAGVYIN--LSKDTFIKVPFKIKBEKEENKPTFDVS-----K	117	
DB	640	EDVLAFEDLTCLKVSTKNGFVTVQSIKGIK-----NKKLEVSLSAED	686	
QY	118	KKDNPQNVHNSQLNSHRKEDLQREHSQKSTQKDTVTATVLDKNNISKSTNN	171	
DB	687	TDDQDEKTDDESSONKSKKKADEHSNTSSSTKN-----DKSNADSKNDSDD	734	

##### RESULT 2

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM

```
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
; FILE REFERENCE: 275601US0
; CURRENT APPLICATION NUMBER: US/11/189,817
; CURRENT FILING DATE: 2005-07-27
; PRIOR APPLICATION NUMBER: 60/598,062
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-189-817-2

Query Match          12.5%; Score 112.5; DB 7; Length 354;
Best Local Similarity 22.4%; Pred. No. 0.06;
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKEFILN-KDTGRVSLKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF 68
Db 153 STTKYAKAKNAKAYKAKNAQKQAVLKKEASS-----YDYL----- 194

QY 69 DGWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKE-----E 108
Db 195 -GWFGGVPPEHKKEENMLSHLYVSKKXENISKENDVDLDEKEEAEETEELKEKNE 253

QY 109 NKPTPDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKST 168
Db 254 ETESEISDEBEKEEKEEENKCKQKEQENNDQKQMEAE-----QNLISKNQ 308

QY 169 TNN 171
Db 309 NNN 311

RESULT 3
US-11-128-660-1
; Sequence 1, Application US/11/128660
; Publication No. US20060024324A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmo
; FILE REFERENCE: 15007dk
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match          12.3%; Score 110.5; DB 7; Length 651;
Best Local Similarity 22.8%; Pred. No. 0.18;
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;

QY 14 KGFILNKDTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63
Db 422 EBAVSEKNAHETVE---HEETVQESNPEKADNDGNVQSNNNELNNEFV-----ESE 472

QY 64 K-----GYQPD-GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIE 102
Db 473 KSEHEARSKAKAEASSYDYTLGWFEFGGVPPEHKKEENMLSHLYVSKKXENISKENDVDLD 532

QY 103 EKKEENKPTFDVSKKQNPQVNHSQLN-----BSHRKEDLQREHSQKSDS 149
Db 533 E-KEEAEETEELKEEKEETESEISDEDEBEKEEKEEKEEKEEKEEKEEKEEKEE 591

QY 150 TKDVATVLDKNNISSKSTNN 171
Db 592 KKQMEA-----QNLISKQNNN 608
```

```
RESULT 4
US-11-087-099-11456
; Sequence 11456, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11456
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Bacillus cereus ATCC 14579
US-11-087-099-11456

Query Match          10.6%; Score 95; DB 7; Length 886;
Best Local Similarity 24.5%; Pred. No. 5;
Matches 49; Conservative 28; Mismatches 75; Indels 48; Gaps 10;

QY 18 LNKDTGEVSELKP-----HRVTVTIQNGKMSSTIVSEEDFILPVY--KGE----- 61
Db 281 INLGANEIFDVRPVEELGKRISIDIQRKIFLNEASVDEEIKIPVYNLKGEPLOQINLQS 340

QY 62 ----LEKGYQPDGWEISG----FEGKIDAGYVINLSKDTPIKPVFKKIEKKEENKPTFD 114
Db 341 EGATLNNG--PIKWNSPGEKIYEFKIDTN--STESKIRFNFGTVIQNIVEKQESQNVILD 396

QY 115 VSKKK--DNPQVNHSQLNESHKEDL-----QREHSQKSDSTKDVTA---TVLDK--- 160
Db 397 KTLQOHINKENLGRNVNAPITKSDLLQIKKGBILKEKNGEIKDITGLEVTNLEKLTLE 456

QY 161 -----NNISSKSTNNPN 173
Db 457 GVGLKNIIEFISNLKQLNNVN 476
```

```
RESULT 5
US-10-485-517-381
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match          10.4%; Score 93.5; DB 6; Length 501;
Best Local Similarity 21.9%; Pred. No. 3.3;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 11 TTVEKFIINKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 66
Db 183 SAITEFQNVQPTNEKMTLDQTKVYVYESVENNESMMDTFVKH-----PIKTGMLNGKKY 237
```

QY 67 QP-----DQWEISGEGKK-----D 81  
Db 238 MVMETNDYKDFWVGQRTISKAKNTRTIIPVVEGKTLVDALVVKVTIDYD 297  
QY 82 AGVINL-SKDTFIKVPFKKIEKKEENKPTFDV-----SKKONPQVNHSQLNESHK 135  
Db 298 GQYHVRVDKEAFTKANTDKSNKGEQDNSAKKEATPATPSKPTSPVEKESQKQDSQKD 357  
QY 136 EDLQ-----REHSQKSDSKQVT-ATVLDKNNISSKSTNNPNK 174  
Db 358 DNKQLPSVEKENDASSEGKDKTPATKPTKGEVSSSTT--PTK 399

RESULT 6

US-10-485-517-244  
; Sequence 244, Application US/10485517  
; Publication No. US2005025629A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; FILE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 244  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-244

Query Match 10.4%; Score 93.5; DB 6; Length 645;  
Best Local Similarity 21.9%; Pred. No. 4.5;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
QY 11 TTVKGFILNKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 66  
Db 327 SAITRFQNPQVPTNEKMTLDQTKYVYVESVENNESWDTFVKH-----PIKTGMLNGKKY 381  
QY 67 QP-----DQWEISGEGKK-----D 81  
Db 382 MVMETNDYKDFWVGQRTISKAKNTRTIIPVVEGKTLVDALVVKVTIDYD 441  
QY 82 AGVINL-SKDTFIKVPFKKIEKKEENKPTFDV-----SKKONPQVNHSQLNESHK 135  
Db 442 GQYHVRVDKEAFTKANTDKSNKGEQDNSAKKEATPATPSKPTSPVEKESQKQDSQKD 501  
QY 136 EDLQ-----REHSQKSDSKQVT-ATVLDKNNISSKSTNNPNK 174  
Db 502 DNKQLPSVEKENDASSEGKDKTPATKPTKGEVSSSTT--PTK 543

RESULT 7

US-11-096-568A-3070  
; Sequence 3070, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 3070  
; LENGTH: 493

; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(493)  
; OTHER INFORMATION: Ceres Seq. ID no. 15172486  
US-11-096-568A-3070

Query Match 10.4%; Score 93; DB 7; Length 493;

Best Local Similarity 26.8%; Pred. No. 3.5;  
Matches 57; Conservative 24; Mismatches 78; Indels 54; Gaps 11;  
QY 3 VVKDPARNTTVKPFIL-----NKDTGE-----VSELKPHR---VTVTIQNGKEM--SSTI 47  
Db 34 VVSDPSVKTTKKKIKIKRVPKKVVGGEASKSLVSEPKKDENGQDSTQSSGKQTADANTI 93  
QY 48 VSEE-----DFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVI----- 86  
Db 94 VTEKKPKGVVPPKKIKITPVSK---KKDETADSNKTETLSDKDEGNVAVVQAQDDTQST 150  
QY 87 ---NLSKDTPIKPVFKIEE--KKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE 141  
Db 151 GKQTANADTTVPEVKTKGVVPPKKQSKTPT---SEKRDN--TADSSKTETKSKDDKKE 205  
QY 142 EHSQKSDSKQVTATVLDKNNISSKSTNNPNK 174  
Db 206 ER-----VTGEKSGAKTDKLGKASDKDQVTVNVKKG 233

RESULT 8

US-11-096-568A-3071  
; Sequence 3071, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 3071  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Glycine max  
; NAME/KEY: misc feature  
; LOCATION: (1)-(493)  
; OTHER INFORMATION: Ceres Seq. ID no. 16625552  
US-11-096-568A-3071

Query Match 10.4%; Score 93; DB 7; Length 493;

Best Local Similarity 26.8%; Pred. No. 3.5;  
Matches 57; Conservative 24; Mismatches 78; Indels 54; Gaps 11;  
QY 3 VVKDPARNTTVKPFIL-----NKDTGE-----VSELKPHR---VTVTIQNGKEM--SSTI 47  
Db 34 VVSDPSVKTTKKKIKIKRVPKKVVGGEASKSLVSEPKKDENGQDSTQSSGKQTADANTI 93  
QY 48 VSEE-----DFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVI----- 86  
Db 94 VTEKKPKGVVPPKKIKITPVSK---KKDETADSNKTETLSDKDEGNVAVVQAQDDTQST 150  
QY 87 ---NLSKDTPIKPVFKIEE--KKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE 141  
Db 151 GKQTANADTTVPEVKTKGVVPPKKQSKTPT---SEKRDN--TADSSKTETKSKDDKKE 205  
QY 142 EHSQKSDSKQVTATVLDKNNISSKSTNNPNK 174  
Db 206 ER-----VTGEKSGAKTDKLGKASDKDQVTVNVKKG 233

RESULT 9





```
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4771
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(140)
; OTHER INFORMATION: Ceres Seq. ID no. 14304111
US-11-096-568A-4771

Query Match
Best Local Similarity 9.7%; Score 87; DB 7; Length 140;
Matches 36; Conservative 28; Mismatches 43; Indels 32; Gaps 7;

QY 40 GKEMSTI-----VSEEDFILPVY--KGELEKGYQFDGWEISGFGKDKAGVYNLSKDT 92
Db 8 GKQASAREDHGSGEEDKIPAYRRGRPQPKMKDD-----FEEEDDEE----- 52

QY 93 FIKPVFKKIEEKKEENKTFDVSKKNPQ-----VNSQLNESHKEDLQREHSQKSD 148
Db 53 --BELVKMEEEEDDSVT---SKKEENRKRKMGNSNTDANKENGLOGSKSRDGS 107

QY 149 STKDVATVLDKNNISSKS 167
Db 108 STKS-TSTGFRNGRRKS 125

RESULT 17
US-11-188-298-1015
; Sequence 1015, Application US/11/188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 1015
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match
Best Local Similarity 9.6%; Score 86; DB 7; Length 443;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 58 YKGELEKGYQFDGWEISGFGKDKAGVYNLSKDTFI----- 94
Db 42 YEEAEDVGVSFDGSSIPGEGIEDSLIFKADPSTVABTPWEGIGRVGYIYKGDPEYQA 101

QY 95 --KPVFKKIEEKKEEN-----KPTFDVSKKD----- 120
Db 102 DPGILKRVLERLEKEGLKAHIGPEPEFYIFKNGTWELHIPDSGGYFDLVGLDKAREIR 161

QY 121 -----NPQVNSQLNESHKEDLQREHSQKSD---STKDVATVLD 159
Db 162 RETALNYPYGLKPEVLHHEVGKAQHEIDFRYDEALRTADNIVSPKHVVKAAR 215

RESULT 18
US-11-196-475-74
; Sequence 74, Application US/11/196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
```

```
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/195,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,336
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match
Best Local Similarity 9.6%; Score 86; DB 7; Length 700;
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;

QY 33 VVTVIQNGKEMSTIVSEEDFILPVYKGELEKGYQFD-----GWEISGFGK 80
Db 258 ITETIENLDQLEKATDDE-----HKETES--QVDAKKQKBEKDKAIDLDKAQKL 309

QY 81 D-AGVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKNPQV-----HSQLN 130
Db 310 DFAEDNLDIQDITVREKLGQENINETNKEKNLPKPGDVSSPKVKOLQIKESLEDLQELK 369

QY 131 ES---HRKEDLQREHSQKSD-----STKDVATVLDKNNISSKSTNNPNK 174
Db 370 ETGDNQKRETEKQIEKIEKSKDEKLLKSKDKDKADGKALDLDR-ELASKASKSEKSK 425

RESULT 19
US-11-072-512-2272
; Sequence 2272, Application US/11/072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
```



```
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2272
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match
Best Local Similarity 9.5%; Score 85.5; DB 7; Length 734;
Matches 39; Conservative 42; Mismatches 58; Indels 41; Gaps 11;

QY 20 KDTG-EVSELKPHRVTV-----TTONKEMSSSTIVSEE--DFILPVYK 59
Db 168 EDTGIYVSRVDPNSIAAKGRIRBGRILQINGEDVQNRBEAVALLSNDECKRIVLLVAR 227
QY 60 GELEKGYQPDGWEISGFKKAGYVYINLSKDTF-----IKPVFKTEE-KCEENKP 111
Db 228 PEIQLD---EGW-----LEDERNE-FLEELNLEMEHEHNEAQPTANEVEQPKQBESEG 279
QY 112 TFDVSKKDNPPQVNHSQLNESHR-KEDLQREHSQKSDSTKDVATVL-DKNNISKSTT 169
Db 280 ITDTATSSN---NHEKDSGVRGRTDSLRNDSSESQENAAEDPNSTSLKSKRDLGQSQDT 336

RESULT 20
US-10-793-626-552
; Sequence 552, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 552
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-552

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 394;
Matches 44; Conservative 13; Mismatches 45; Indels 44; Gaps 10;

QY 10 NTTVKEFILNKDTGEVSELKPHRVTVTTONKEMSTI---VSEEDFILPVYKGELEKGY 66
Db 55 NINVSQFENPKTHE-SEV--HTASSRANNGRFPSPVPFKLASEHIDITQVKG----- 105
QY 67 QPDGWEISGPEG---KKDAGYVYINLSKDTFIKPVFKIEBKKEENKPTFDVSKKKNPQ 123
Db 106 -----TGFEGRVTKDQNTQINNPD-----QBEKEPKQT-----DKKH-S 142
QY 124 VNHQQLNESHKEDLQREHSQKSD 148
Db 143 TNHCDPLHQSSTK-----NEHSPLSN 163

RESULT 21
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 472;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 101 IEHKKEENKPTFDVSKKKNPPQVNHSQLNESHKEDLQ-----REHSQKSDS--- 149
Db 3 MEENKQPNKE--NMSNKDDNA-----THLNDSHRNDELLEFRNKNAQRERRRRIDNOSK 56
QY 150 TKDVTAT-----VLDKNNISSKSTTNPKNK 174
Db 57 EKDATSTQSLKETPKMDKPLDNHKS--HNQNK 86

RESULT 22
US-10-485-517-170
; Sequence 170, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 170
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-170

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 636;
Matches 35; Conservative 35; Mismatches 61; Indels 28; Gaps 7;

QY 17 ILNKDTGEVSELKPHRVTVTTONKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGF 76
Db 27 IVTKDYSKQSVNAGSKNGTLDLSRYLNSALYYLEDYI--IYAILGTLTKRYE-----GD 78
QY 77 EGKGAGYVYINLSKDTFIKPVFKK-----IEBKCEE-ENKPTFDVSKKKNPPQVNHSQLN- 130
Db 79 NIYKEA-----KDRLEKVLREDOYLLERKKSQYEDYKQYANYKKENFRTDLKMANF 131
QY 131 ESHRKEDLQREHSQKSDSTKDV-----TATVLDKNN 162
Db 132 HKYNLEELSMKEYNELQALKRALDDPHREVKDIDKNS 170

RESULT 23
US-11-196-475-66
; Sequence 66, Application US/11196475
```



```
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(974)
; OTHER INFORMATION: Ceres Seq. ID no. 13623377
US-11-096-568A-26839

Query Match
Best Local Similarity 9.4%; Score 84.5; DB 7; Length 974;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPKPARNT-----TVKEFILNKDTGVSSELKPHRVTVTIQNGKEMSTIVS 49
Db 791 IVKDFPKDVLRIIDSPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 843
QY 50 BEDFILPVYKGELEKGYQDFGWEISGFEG-----KDGAGYVINLSKDTFIKPVFK 100
Db 844 DP-----QKPIDGGWEFLNMEASDSETEDETSQGVV-----PSDAE 882
QY 101 IEKKKEENKPTFDVSKKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 883 PESESDDDSDSESLVESDDDESDSEDESEKGTWEELERASNADREHGAESDSSEE 942

RESULT 27
US-11-096-568A-26838
; Sequence 26838, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26838
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1055)
; OTHER INFORMATION: Ceres Seq. ID no. 13623376
US-11-096-568A-26838

Query Match
Best Local Similarity 9.4%; Score 84.5; DB 7; Length 1055;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPKPARNT-----TVKEFILNKDTGVSSELKPHRVTVTIQNGKEMSTIVS 49
Db 872 IVKDFPKDVLRIIDSPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 924
QY 50 BEDFILPVYKGELEKGYQDFGWEISGFEG-----KDGAGYVINLSKDTFIKPVFK 100
Db 925 DP-----QKPIDGGWEFLNMEASDSETEDETSQGVV-----PSDAE 963
QY 101 IEKKKEENKPTFDVSKKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 964 PESESDDDSDSESLVESDDDESDSEDESEKGTWEELERASNADREHGAESDSSEE 1023

RESULT 28
US-11-096-568A-26837
; Sequence 26837, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
```

```
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26837
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1057)
; OTHER INFORMATION: Ceres Seq. ID no. 13623375
US-11-096-568A-26837

Query Match
Best Local Similarity 9.4%; Score 84.5; DB 7; Length 1057;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPKPARNT-----TVKEFILNKDTGVSSELKPHRVTVTIQNGKEMSTIVS 49
Db 874 IVKDFPKDVLRIIDSPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 926
QY 50 BEDFILPVYKGELEKGYQDFGWEISGFEG-----KDGAGYVINLSKDTFIKPVFK 100
Db 927 DP-----QKPIDGGWEFLNMEASDSETEDETSQGVV-----PSDAE 965
QY 101 IEKKKEENKPTFDVSKKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 966 PESESDDDSDSESLVESDDDESDSEDESEKGTWEELERASNADREHGAESDSSEE 1025

RESULT 29
US-11-191-374-16
; Sequence 16, Application US/11191374
; Publication No. US20050260673A1
; GENERAL INFORMATION:
; APPLICANT: Hresko, Michelle Coutu
; APPLICANT: McLaird, Merry B.
; APPLICANT: Williams, Deryck J.
; APPLICANT: Frevert, Anita M.
; APPLICANT: Chiapelli, Brandi
; APPLICANT: Baublite, Catherine
; APPLICANT: Kloek, Andrew P.
; APPLICANT: Davila-Aponte, Jennifer A.
; APPLICANT: Bradley, John D.
; APPLICANT: Xu, Siqun
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE
; FILE REFERENCE: 12557-015001
; CURRENT APPLICATION NUMBER: US/11/191,374
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/771,708
; PRIOR FILING DATE: 2004-02-04
; PRIOR APPLICATION NUMBER: US 60/444,771
; PRIOR FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-191-374-16

Query Match
Best Local Similarity 9.4%; Score 84.5; DB 7; Length 1065;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARTTVKEFILNKDTG-----EVSSELKPHRVTVTIQNGKEMSTIVSEE----- 51
Db 570 DFKNLCVTKEI---DTGASAAANVPETK-HRVSGTVVVEGDKSKSOLLATKKVKKPTIKN 625
QY 52 -----DFILPV-----YKG-----ELEKGYQDFGWEISGFEGKKD----- 81
Db 626 TEHRRAPSTVFIPGPVEVKAEAIQTICNYEIGIKVQINNGEPFSG--VIFVKNKFTCRV 683
```



Query Match 9.4%; Score 84.5; DB 7; Length 1069;  
Best Local Similarity 23.0%; Pred. No. 46;  
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNVTKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51  
DB 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKPTIKN 629  
QY 52 -----DFILPV-----YKG---ELEKGYQFDGWEISGPEGKKD----- 81  
DB 630 TEHRRAPSTVPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687  
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKEEENKPTFDVSKKKNPQVNHSQLNESH 133  
DB 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171  
DB 733 LKDEI---EFPQRQAARFDCGLVDLLNGTYKSTVVIQTNN 772

RESULT 33  
US-11-191-375-17  
; Sequence 17, Application US/11191375  
; Publication No. US20050260674A1  
; GENERAL INFORMATION:  
; APPLICANT: Hresko, Michelle Coutu  
; APPLICANT: McLaird, Merry B.  
; APPLICANT: Williams, Deryck J.  
; APPLICANT: Frevert, Anita M.  
; APPLICANT: Chiapelli, Brandi  
; APPLICANT: Baublite, Catherine  
; APPLICANT: Kloek, Andrew P.  
; APPLICANT: Davila-Aponte, Jennifer A.  
; APPLICANT: Bradley, John D.  
; APPLICANT: Xu, Siqun  
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE  
; FILE REFERENCE: 12557-015001  
; CURRENT APPLICATION NUMBER: US/11/191,375  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US/10/771,708  
; PRIOR FILING DATE: 2004-02-04  
; PRIOR APPLICATION NUMBER: US 60/444,771  
; PRIOR FILING DATE: 2003-02-04  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 1069  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-11-191-375-17

Query Match 9.4%; Score 84.5; DB 7; Length 1069;  
Best Local Similarity 23.0%; Pred. No. 46;  
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNVTKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51  
DB 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKPTIKN 629  
QY 52 -----DFILPV-----YKG---ELEKGYQFDGWEISGPEGKKD----- 81  
DB 630 TEHRRAPSTVPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687  
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKEEENKPTFDVSKKKNPQVNHSQLNESH 133  
DB 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171  
DB 733 LKDEI---EFPQRQAARFDCGLVDLLNGTYKSTVVIQTNN 772

RESULT 34  
US-11-191-588-17  
; Sequence 17, Application US/11191588  
; Publication No. US20050282222A1  
; GENERAL INFORMATION:  
; APPLICANT: Hresko, Michelle Coutu  
; APPLICANT: McLaird, Merry B.  
; APPLICANT: Williams, Deryck J.  
; APPLICANT: Frevert, Anita M.  
; APPLICANT: Chiapelli, Brandi  
; APPLICANT: Baublite, Catherine  
; APPLICANT: Kloek, Andrew P.  
; APPLICANT: Davila-Aponte, Jennifer A.  
; APPLICANT: Bradley, John D.  
; APPLICANT: Xu, Siqun  
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE  
; FILE REFERENCE: 12557-015001  
; CURRENT APPLICATION NUMBER: US/11/191,588  
; CURRENT FILING DATE: 2005-07-28  
; PRIOR APPLICATION NUMBER: US/10/771,708  
; PRIOR FILING DATE: 2004-02-04  
; PRIOR APPLICATION NUMBER: US 60/444,771  
; PRIOR FILING DATE: 2003-02-04  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 1069  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-11-191-588-17

Query Match 9.4%; Score 84.5; DB 7; Length 1069;  
Best Local Similarity 23.0%; Pred. No. 46;  
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNVTKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51  
DB 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKPTIKN 629  
QY 52 -----DFILPV-----YKG---ELEKGYQFDGWEISGPEGKKD----- 81  
DB 630 TEHRRAPSTVPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687  
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKEEENKPTFDVSKKKNPQVNHSQLNESH 133  
DB 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171  
DB 733 LKDEI---EFPQRQAARFDCGLVDLLNGTYKSTVVIQTNN 772

RESULT 35  
US-11-196-400-3  
; Sequence 3, Application US/11196400  
; Publication No. US20050287166A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 200773USODIV  
; CURRENT APPLICATION NUMBER: US/11/196,400  
; CURRENT FILING DATE: 2005-08-04  
; PRIOR APPLICATION NUMBER: US/09/742,096  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 08/973,462  
; PRIOR FILING DATE: 1998-02-06  
; PRIOR APPLICATION NUMBER: PCT/FR96/00894  
; PRIOR FILING DATE: 1996-06-12  
; PRIOR APPLICATION NUMBER: FR 95/07007







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.  
OM protein - protein search, using sw model  
Run on: April 24, 2006, 14:50:52 ; Search time 18.2953 Seconds  
(without alignments)  
915.083 Million cell updates/sec  
Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTVTKSEFILNK.....ATVLDKNNISSKSTNNPNK 174  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR 80:\*\*  
1: pirl:\*\*  
2: pirl:\*\*  
3: pirl:\*\*  
4: pirl:\*\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2140	2 F95074	serine proteinase,
2	894	99.7	2144	2 A97942	metalloproteinase,
3	114.5	12.8	1038	2 JC5497	claustrin - chicke
4	110	12.3	558	2 T18467	hypothetical prote
5	109	12.2	665	2 B71609	hypothetical prote
6	104	11.6	540	2 D86432	hypothetical prote
7	103.5	11.5	325	2 T18283	hypothetical prote
8	102.5	11.4	385	2 T20410	hypothetical prote
9	102	11.4	622	2 A90570	lipoprotein (impor
10	101.5	11.3	312	2 G81339	probable membrane
11	101.5	11.3	535	2 T37189	hypothetical prote
12	101	11.3	700	2 S67610	probable membrane
13	100	11.1	219	2 B72291	hypothetical prote
14	100	11.1	1875	2 S38173	myosin-like protei
15	99	11.0	211	2 T25911	hypothetical prote
16	99	11.0	1397	2 T10466	DNA topoisomerase
17	97.5	10.9	614	2 A84152	hypothetical prote
18	97.5	10.9	1345	2 S46817	hypothetical prote
19	97	10.8	644	2 T47835	hypothetical prote
20	97	10.8	988	2 T14188	hypothetical prote
21	96.5	10.8	456	2 T05612	hypothetical prote
22	96.5	10.8	867	2 T27136	hypothetical prote
23	96.5	10.8	871	2 T27135	hypothetical prote
24	96	10.7	491	2 C97267	hypothetical prote
25	96	10.7	1939	2 T18372	repeat organellar
26	96	10.7	2500	2 G71609	hypothetical prote
27	95.5	10.6	3724	2 T18427	hypothetical prote
28	95	10.6	348	2 T37271	cylicin II - human
29	95	10.6	1202	1 S05362	probable DNA-dirc

30	94.5	10.5	277	2 D70214	surface lipoprotei
31	94.5	10.5	670	2 T28391	ORF MSV230 hypothe
32	94	10.5	210	2 T28771	hypothetical prote
33	94	10.5	553	2 T15094	hypothetical prote
34	94	10.5	947	2 T03795	ornithine decarbox
35	93.5	10.4	456	2 T03045	hypothetical prote
36	93.5	10.4	645	2 E89883	conserved hypotet
37	93.5	10.4	649	2 S42488	dnak-type molecula
38	93	10.4	629	2 G96542	hypothetical prote
39	93	10.4	2464	1 QRMSP1	microtubule-associ
40	92.5	10.3	443	2 S66040	serine-type D-Ala-
41	92.5	10.3	571	2 B64469	hypothetical prote
42	92.5	10.3	1702	2 A41859	Iga-specific metal
43	92.5	10.3	1888	2 T39009	hypothetical prote
44	92.5	10.3	1997	2 F71607	DNA helicase II BR
45	92	10.3	688	2 A47705	triacylglycerol li

ALIGNMENTS

RESULT 1  
P95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: P95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: P95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match 100.0%; Score 897; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-56;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTVTKSEFILNKDTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1933 KIVVKDFARNTVTKSEFILNKDTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKG 1992  
QY 61 ELEKGYQFDGWEISGFEKGKDGAGVYVNLKDTFKVPFKIEKKKEENKPTFDVSKKDD 120  
DB 1993 ELEKGYQFDGWEISGFEKGKDGAGVYVNLKDTFKVPFKIEKKKEENKPTFDVSKKDD 2052  
QY 121 NPQVNHSQLNESHKREDIQRHEHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 174  
DB 2053 NPQVNHSQLNESHKREDIQRHEHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2106

RESULT 2  
A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
C:Cross-references: UNIPROT:Q8DQP7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C:Genetics:  
A:Gene: pta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.7%; Score 894; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 4.1e-56;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARTTVKPEILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1937 KIVVKDPARTTVKPEILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 1996

QY 61 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKVPFKKIEEKKKEENKPTFDVSKKKD 120  
DB 1997 ELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKVPFKKIEEKKKEENKPTFDVSKKKD 2056

QY 121 NQVNHISQLNESHKEDLQREHSHQSDSTKDTVTATVLDKNNISSKSTNNPNK 174  
DB 2057 NQVNHISQLNESHKEDLQREHSHQSDSTKDTVTATVLDKNNISSKSTNNPNK 2110

RESULT 3  
JC5497  
Clausstrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: JC5497; PC4334; S37561  
R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Clausstrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BURL>  
A:Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:g406318; PID  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83/299-412/485-502 <BUR2>  
A:Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BPF4; UNIPARC:UPI000017BPF5  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Aen) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 12.8%; Score 114.5; DB 2; Length 1038;  
Best Local Similarity 23.5%; Pred. No. 1.4;  
Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;

QY 2 IVVKDPARN---TTVKEFILNKDTGEVSELKPHRVTVTIQ----- 39  
DB 514 VTQKDLTGNIASPAVKQAKLKORTDSKSLKPAKTTTKDCCQKNLKKKHSLSLQVQQ 573

QY 40 -----GKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGK 79  
DB 574 LEKPKLESKEKTPVKKKAVKPEKTIIVAEKDV-----TTKEEQLGKSETSEKQAS 625

QY 80 KAGYVYVNLKDTFIKVPFKKIEEKKKEENKPTFDVSKKKDNPQVNHISQLNESHKEDL 138  
DB 626 EKQDVKPKVTKEKSVKKEKVPKPEKCKEKEKPKKEVSKKEEKLPI---KKEEKPKCKEDI 682

QY 139 QRE-----EHSQKSDSTKDV 153  
DB 683 KKEVKKKEVKKKEKKEAKEV 702

RESULT 4  
TI8467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: TI8467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: TI8467  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465C

Query Match 12.3%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 1.5;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 47 IVSEEDFILPVY-----KGELEKGYQPDGWEISGPEGKK---DAGYVYVNLKDTFIKVP 97  
DB 60 ILGPEDDILYEYCISQLKQSKKE--KADGEEDKYLNAKKLAKINLTGPIGNKKSDIFIEEL 117

QY 98 PKKI--BEKKER-----ENKPTFDVSK-KQNPQVNHISQLNE-----SHRK 135  
DB 118 LELLTNEKKEEHIADTLNENK-TNDIKKVNENINENYVYVNLKDTFIKVP 176

QY 136 E-----DLQREH-----SQKSDSTK---DVTATVLDKNNISSKSTTN 170  
DB 177 EHNINNVNLKKEKEYTDIQRDKRKHRSLSQKSDSYKRPFNKRTKSTIER-SLSNKRDE 235

QY 171 NPNK 174  
DB 236 KTNK 239

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; GB:AE001362; NII  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680w

Query Match 12.2%; Score 109; DB 2; Length 665;  
Best Local Similarity 26.1%; Pred. No. 2.1;  
Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KDFARNTTVKPEILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELE 63  
DB 114 KDDNNNNNGTKQIEKKNKINKSDL--HRQNELNLQSGK-----NEQDI-----NKNK 158

QY 64 KGYQPDGWEI--SGPEGKDGAGYVNLKDTFIKVPFKKIEEKK-----EENKP 111  
DB 159 KGKQ-----DISNSAENKKD-----VKEGVKELEKKEKKEKESKIDDKHVEENKK 202

QY 112 TFD-----VSKKKNPQVNHISQLNESHKEDLQR--EEHSQKSDSTKDTVTATVLDKNNISSK 166

Db 203 SDHKVEENKSDDHKVEENKSDDHKBEVKKVEHEDEE-----DKKEKSE 253

QY 167 STTNPNK 174

Db 254 NKNDENK 261

RESULT 6

D86432

hypothetical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: D86432

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.P.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 <STO>

A:Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI000000ABP35; GB:AE005172; NID:g4587525; PMID:11130712

C:Genetics:

A:Map position: 1

Query Match 11.6%; Score 104; DB 2; Length 540;

Best Local Similarity 25.3%; Pred. No. 3.8;

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TTVKFELNKDTGVESELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 70

Db 27 TTVKA-VVEETKVEDESKP-----EGVEKSASFKEESDFPADLKESEKK----- 70

QY 71 WISGPEGKDGAGVYINLSDTFIKVPFK-----IEBKKEENKPTFDVSKKKNQVNH 127

Db 71 -ALSCLKLBEAIVDN-----TLTKKESSPMKKEEVVYKPEAEVEKKKE--EAAEE 123

QY 128 QLNESHKEDLQREHSOKSDSTKDTATVLDKNISS 165

Db 124 KVEEKKSEAVTETAPKATVEAVVTEIIPKEVTT 161

RESULT 7

T18283

hypothetical protein G5 - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T18283

R:Riebel Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh Genetics 148, 1117-1125, 1998

A:Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1

A:Reference number: Z14684; MUID:98198836; PMID:9539429

A:Accession: T18283

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-325 <RIE>

A:Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; PMID:9539429

C:Genetics:

A:Introns: 85/1

Query Match 11.5%; Score 103.5; DB 2; Length 325;

Best Local Similarity 23.8%; Pred. No. 2.3;

Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 31 HRVTVTIQNGKMSSTIVSEEDFILPVYK-GEL--EKGYQPDGWEISGPEGK----- 79

Db 57 HRTITSIKN--RFSVKKIGDEKLPFRISKNGELIVLNEFDNPHK--EGHLEKSKMF 112

QY 80 ---KDAGYV-----INLSKDTFIKPV-----FKK----- 100

Db 113 NHKDSGYATNNEIEIIFLESTLCKEITATQKNSYKRNINIKLPPEEEEEEEEEEE 172

QY 101 --IEBKKEENKPTFDVSKKKNQVNHSQLNESHKKE-----DLQREHSOKSDSTK 151

Db 173 EEEEEQEVEKPTTISEEEEEETPAVSEEEEEEEEEETPAVSEEEEEKEEEEEQEEDKEK 232

QY 152 D 152

Db 233 D 233

RESULT 8

T20410

hypothetical protein E02A10.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T20410

R:Thomas, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19271

A:Accession: T20410

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-385 <WIL>

A:Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z81053; PIDN:CAB02877.1

A:Experimental source: clone B02A10

C:Genetics:

A:Gene: CESP:E02A10.2

A:Map position: 5

A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 11.4%; Score 102.5; DB 2; Length 385;

Best Local Similarity 26.2%; Pred. No. 3.4;

Matches 48; Conservative 31; Mismatches 55; Indels 49; Gaps 9;

QY 1 KIVVKDPARNT-----TVKEFILNKD-----TGEVSELKPHRVTVTIQ- 38

Db 215 RIILNENTKETAASSIKTLKKEKVAGQDIYVCKEKPAPFTATDDFCSLQKENVHCTILR 274

QY 39 -NGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGVYINLSKDTFIKPV 97

Db 275 INHKEVAEK--NEEDKKEEPKKEEKEVEKKE--EDEKDE-----EP- 316

QY 98 PKKIEBKKEENKPTFDVSKKKNQVNHSQLNESHKEDLQREHSOKSDSTKDVTTATV 157

Db 317 -KGEKKEEKEKKE--EVEKKEE-----EEKKEBEPKKEEKEEKEEKEEVEBSEKV 368

QY 158 LDK 160

Db 369 BEK 371

RESULT 9

A90570

lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004

C:Accession: A90570

R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu.

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: A90570

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-622 <KUR>

A:Cross-references: UNIPROT:Q980A1; UNIPARC:UPI00000C80A1; GB:AL445566; PID:g14089879;

A:Experimental source: strain UAB CTIP

C;Genetics:  
A;Gene: MYPU 4650  
A;Genetic code: SGC3

Query Match 11.4%; Score 102; DB 2; Length 622;  
Best Local Similarity 24.9%; Pred. No. 6.3;  
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;  
  
QY 9 RNTTVEKPTL-NKOTGEVSEL-----KPHRV--TWTIQNG--KEM--SSTI--V 48  
DB 26 KNDKSAKFNKSSDLISISEKQFQFNKNDLDRNQKLIIEGVITFINSETKBIIVKKEITLNL 85  
QY 49 SBEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGVYVNLKSDTPIKVPFKKIEBKKEE 108  
DB 86 SEDNIFSLNLNNAKQFQDDEP-----VSKDEKFKIQEINFSQYEQ 129  
QY 109 NKPTFDVSKKD-----NPQVNHSQLNES--HRKEDLQREHSOKSDSTKVDTATVLDKRN- 161  
DB 130 -KITDNISSKEDEKKNPKDNNSNNSDQKNDLQKNNSDKLNDVQDEKANKENSNS 188  
QY 162 NISSKSTTNPNK 174  
DB 189 NDSKEKNDENTNK 201

## RESULT 10

probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81339  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10689204  
A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL139076; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 11.3%; Score 101.5; DB 2; Length 312;  
Best Local Similarity 25.0%; Pred. No. 3.1;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;  
  
QY 15 EFTLNKDTGEVSELKPHRVTTVTIQNKEMSSNI---VSEEDFILPVYK-----GELE 63  
DB 5 DFIKERQNTQRQKWLKFSR---AINQKPLDDDLDRDEISSDDILRRFPKKTTPNKPLEELD 61  
QY 64 KGQYQDGEISGPEGKDGAGVYVNL---SKDTPIKVPFKKIEBKKEENKPT---FDVSK 117  
DB 62 EYESKHTKNSIYLKED---LINVKLEKQSLAKKIFSKMKERRKEENKTKKNFLFSR 118  
QY 118 KQNP-----QVNHSQLNESHRKEDLQREHSOKSDSTKV--TATVLDKRNISSK--STT 169  
DB 119 KKANEIKNTQTQIQTKSNQATTQTKQEKELTNSIEKIQTETTKIQKPLIEKLDVYK 178  
QY 170 NNPN 173  
DB 179 NQPN 182

## RESULT 11

T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37189  
R;Leimbac, D.; Minx, M.

submitted to the EMBL Data Library, February 1996  
A;Description: The sequence of C. elegans cosmid C02H7.  
A;Reference number: Z20523

A;Accession: T37189  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-535 <LRI>  
A;Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:U49945; PIDN:AAC47924.1;  
A;Experimental source: strain Bristol N2; clone C02H7  
C;Genetics:  
A;Gene: CESP:C02H7.1  
A;Map position: X  
A;Introns: 47/3, 100/3, 149/3, 304/2, 347/3, 458/3

Query Match 11.3%; Score 101.5; DB 2; Length 535;  
Best Local Similarity 21.0%; Pred. No. 5.7;  
Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;  
  
QY 4 VKDFARNTTVKBEFILNK-----DTGHSVSELKPHRVTTVTIQNKEMSSSTIVSEEDFILPVY 58  
DB 55 IKSAGNDKNTKTAFLDKLILKLDGSLKNVKAAIL---SGKDAEET----- 98  
QY 59 KGELEKGYQFDGWEISGPEGKDGAGVYVNLKSDTPIKVPFKKIEBKKEENKPTFDVSKK 118  
DB 99 ----NQMLQMLGTNATSFNSRNGTG-----BEKKKKKKVKKEDKGDDEEKST---TKK 144  
QY 119 KDNQVNHSQLNESHRKEDLQREHSOKSDSTK-----DVTATVLDKNNISSKST 168  
DB 145 RSSKKEETHEKEKKKSAEKEKKKKKKSSSKSRHKHSKSSSEKSKSEKKEKKEKST 204  
QY 169 TNNPNK 174  
DB 205 TDEKPK 210

## RESULT 12

S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein D2483  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S67610  
R;Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S67608  
A;Accession: S67610  
A;Molecule type: DNA  
A;Residues: 1-700 <WAM>  
A;Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069BFF; EMBL:Z74122; NID:gl431087; P1  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:BRE1; MIPS:YDL074C  
A;Cross-references: SGD:S0002232  
A;Map position: 4L  
C;Keywords: transmembrane protein  
P;69-85/Domain: transmembrane #status predicted <TM>

Query Match 11.3%; Score 101; DB 2; Length 700;  
Best Local Similarity 22.4%; Pred. No. 8.5;  
Matches 46; Conservative 39; Mismatches 60; Indels 60; Gaps 10;  
  
QY 5 KDFARNTTVKBEFILNK-----DTGEVSELKPHRVTTVIQ-----NGKEMEST 46  
DB 379 KDVVIRIARTDLDLSKTALEAKSKTEVLSDLQ-HAIDILKEQWTKIDQSRNDTKSSST 437  
QY 47 IVSEEDFILPVYKGELEKGYQFDGWEISGPEGK-----DAGYV 85  
DB 438 ----QDALIKEIQ-DLEKGR----ELSDLTHKKYSEIINHESVLSKLTVEKTKADQKYF 488  
QY 86 INL-SKDTF---IKPVFKKIEBKKEENKPTFDVSKKDNQVNHSQLNESHRKEDLQRE 141  
DB 489 AAMRSKSILIEIKTLTKSLSKSNEL-----ILQKDSRLLQOKIGNLHQLDLSQN 541

[illegible]

A:Molecule type: DNA  
A:Residues: 1-1397 <CH>  
A:Cross-references: UNIPARC:UPI00000006D7; EMBL:X79345; NID:g994807  
C:Genetics:  
A:Gene: TopoII  
A:Map position: 14  
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.0%; Score 99; DB 2; Length 1397;  
Best Local Similarity 22.2%; Pred. No. 26;  
Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVVKDFAR-----NTTVKEPILNKDCTGEVSELKPHRVTVTIQNGKEMSSSTIVSEE 51  
DB 1092 KVLVELYRKGVDYPKDINKIKKEBIFEOELLDAAD-NPE-----DNEEIIAGITVKDY 1144  
QY 52 DFIL--PVYKGBLEKGYQPDGWEISGFEGKDGAGVINLSKDTFIKPVFKIEE-----103  
DB 1145 DYLLSNPIFSLTLEK---VEDLLTQLKEKERLEILRNITVETMWLKDIEKVEEAIEFOR 1201  
QY 104 -----KKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQ---KSDSTKDVTA 155  
DB 1202 NVLSNREESNK--FKVARKQ-----GPSSMKKKKKKKLSSDESEGGDTSDSSEFLVN 1254  
QY 156 TVLDKNNISSKSTNNPN 173  
DB 1255 TLNKKNTWKTKITSSNN 1272

RESULT 17  
A84152  
hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: A84152  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11059132  
A:Accession: A84152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <STO>  
A:Cross-references: UNIPROT:Q9K5S1; UNIPARC:UPI0000004396; GB:AP001520; GB:BA0000004; NID  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH4017

Query Match 10.9%; Score 97.5; DB 2; Length 614;  
Best Local Similarity 27.3%; Pred. No. 13;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 26 SELKPHRVTVTIQNGKEMSSSTIVSE-----EDFILPVYKGBLEKGYQPDGW-----EISG 75  
DB 386 SEPKPE--TYTLQTAIQM-TPIVNEYSPTREEFL-----ARKAHQLDGHADVSKVTY 435  
QY 76 FEGKDGAGVINLSKDTFIKPVFKIEEK--EBENKPTFDVSKKK---DNPQVNHSQL 129  
DB 436 FAGRNIASQL-----GKIEEKLQDKYNNYTFDFPKKEVNVVQSPFKSTSA 482  
QY 130 NESHKEDLQREHSOKSDSTKDVTAIVLD 159  
DB 483 GKVGIGVLDGRKYHKGEBLERLSKSQIE 512

RESULT 18  
S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C:Accession: S46817  
R:Favello, T.

submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46817  
A:Molecule type: DNA  
A:Residues: 1-1345 <PAV>  
A:Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:g500825; PID  
C:Genetics:  
A:Cross-references: SGD:S0001122  
A:Map position: 8R  
C:Superfamily: uncharacterized conserved protein  
C:Keywords: transmembrane protein

Query Match 10.9%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 32;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKGYQPDGWEISGFEGKK-DAGVYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSK 117  
DB 1109 KGAIEKG-----SVEGQKVSQYMLSELRII-----SRAKSKPVKVKWK 1149  
QY 118 KQNPQVNHSQLNESHKEDLQREHSOKSDSTKDVTAIVLD--KNNISSKSTTN---NP 172  
DB 1150 SHDKHRPPHSHKVE-----QKSSERKSDNDKDLTHILDFVQNNFSSIFMVKLLSP 1201  
QY 173 NK 174  
DB 1202 QK 1203

RESULT 19  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NYA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 10.8%; Score 97; DB 2; Length 644;  
Best Local Similarity 22.9%; Pred. No. 15;  
Matches 43; Conservative 29; Mismatches 72; Indels 44; Gaps 7;

QY 8 ARNTTVKEFIKND--TGEV--SELKPHRVTVT-----IQNGKEMSSSTIVSEE--- 51  
DB 476 SRRSKEIAVAVAKDTKGRAGNNIKQDTDTKTESDDDDDEKEENSKTEKKTADVADKKSV 535  
QY 52 -DFILPVYKGBLEKGYQPDGWEISGFEGKDGAGVINLSKDTFIKPVFKIEEKKKEENK 110  
DB 536 ADFLKRIKKNSPQKKG-----ETTSXNQKQKDGNV-----KKENDHQ 572  
QY 111 PTFVSKKONPQVNHSQLNESHKEDLQREHSOKSDS---TKDVTATVLDKNNISSK 166  
DB 573 KKSQGNVKKSKVKPRELRSSTGRKRVENNNNSKSSKRRKQTKETAIVATGKGRGSG 632  
QY 167 STNNPNK 174  
DB 633 KDDKQPRK 640

RESULT 20  
T14188



hypothetical protein T28D5.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, A.; et al. 1999  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z17931  
A:Accession: T14188  
A:Molecule type: DNA  
A:Residues: 1-988 <BEV>  
A:Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI00000A3618; EMBL:AL109819  
A:Experimental source: cultivar Columbia; BAC clone T28D5  
C:Genetics:  
A:Gene: ATSP-T28D5.30  
A:Map position: 4  
A:Introns: 162/3; 201/3; 416/3; 438/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C:Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 10.8%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 24;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

QY 39 NGKEMSTIVSEDFILPVYKGELEKGY-----QPDGWEISGPEGKDGAGYVNLNLSKDTF 93  
DB 383 NGRQNSNVQSSVDEILSYTKVPSGVGLNVSEIDIVELDDVRSAGGLSPVQRDN- 441  
QY 94 IKPVFKKIBEKKE-----ENKPTFDVSKKKNPQVNHSQLNESHKREDLQREE 142  
DB 442 VEPVGDDVRSSGDMSPNAAANVRGEPATFDIMESDNPGRDNVAPMEDHRSVQLSP 501  
QY 143 HSQKSDSTKDT--ATVLDKNNISSKSTNNPNK 174  
DB 502 HVL--GAKDVTDSPTDKGVNDVTDASDPT 532

RESULT 21  
T05612  
hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.; et al. 1999  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81F5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 10.8%; Score 96.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 11;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 20 KDTGEVSEKLP-HRYVTVTQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGREG 78  
DB 245 KKKGKODPLKPKHPVSAFLVYANERRAALREENKSVVEAK-----ITGEEWKNLSD 296  
QY 79 KIDAGY--VINLSKDTFIKVPVKITEKKEEENKPTFDVSKKKNPQVNHSQLNESHKRE 136  
DB 297 KKKAPYKAKKKETYLQ-AMEEYKTYKEE-----ALSQKG-----EEELKKHKQE 345  
QY 137 DLQREHSQKSDTKDVTATVLDKNNISSKSTNN--PNK 174  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKNENVPNK 378

RESULT 22  
T27136

hypothetical protein Y53C12B.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27136  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27136  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-867 <WIL>  
A:Cross-references: UNIPROT:O18218; UNIPARC:UPI0000075065; EMBL:Z99278; PIDN:CAB16493.  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3a  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 714/2; 864/3

Query Match 10.8%; Score 96.5; DB 2; Length 867;  
Best Local Similarity 25.0%; Pred. No. 23;  
Matches 45; Conservative 33; Mismatches 63; Indels 39; Gaps 8;

QY 9 RNTTVKEFILNKDTG-----EVSELKPHRYVTVTQNGKMSSTIVSEDFILPVYKGE 62  
DB 651 KKTPKCKDPKVDQEAISIEATKTEISEENPK--TDDIQSKDDVTS-----KSEL 695  
QY 63 E----KGYPQDQWEISGPEGKDGAGYVNLNLSKDTFIKVPVKIE---EKKEEENKPTFD 114  
DB 696 HCYRCHYQLPABEVSSHNIRKNG-----DLWCEHM-KKIKGCHCEATGEGHHPLI 747  
QY 115 VSKKKNPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
DB 748 CPKKEEERVAKSR--ESSQKPIDPQEISDDQDVTDPDQIVQDNQSHKSRNSNR 805

RESULT 23  
T27135  
hypothetical protein Y53C12B.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27135  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27135  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-871 <WIL>  
A:Cross-references: UNIPROT:O62486; UNIPARC:UPI0000079598; EMBL:Z99278; PIDN:CAB16492.1  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3b  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

Query Match 10.8%; Score 96.5; DB 2; Length 871;  
Best Local Similarity 25.0%; Pred. No. 23;  
Matches 45; Conservative 33; Mismatches 63; Indels 39; Gaps 8;

QY 9 RNTTVKEFILNKDTG-----EVSELKPHRYVTVTQNGKMSSTIVSEDFILPVYKGE 62  
DB 654 KKTPKCKDPKVDQEAISIEATKTEISEENPK--TDDIQSKDDVTS-----KSEL 698  
QY 63 E----KGYPQDQWEISGPEGKDGAGYVNLNLSKDTFIKVPVKIE---EKKEEENKPTFD 114  
DB 699 HCYRCHYQLPABEVSSHNIRKNG-----DLWCEHM-KKIKGCHCEATGEGHHPLI 750  
QY 115 VSKKKNPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
DB 751 CPKKEEERVAKSR--ESSQKPIDPQEISDDQDVTDPDQIVQDNQSHKSRNSNR 808

RESULT 24



[illegible][illegible]



Query Match 10.4%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 20.3%; Pred. No. 18;  
Matches 41; Conservative 39; Mismatches 83; Indels 39; Gaps 5;  
  
QY 3 VVKDPAINTVTVKEFILNKDTGVSSELKPHRVTVTTIONGKEMSTIVSEEDFILPVYKGL 62  
DB 70 VVQSSISNTPITEI--KUTPYRE--TPIKRTITVNTVKTSSSISGNGRNLDDDL 126  
QY 63 EKGYPQDWEISGFEGKDDAGVYVNLSDTFI-----RPVFKKIBEKKE-----BENKPT 112  
DB 127 PDDRYKSPTRKFGQEKDEDIRLIPKSNIGSSKYPVLFRVEENENKKIHIQKES 186  
QY 113 FVSKKKQNPQ-----VNHSQLNESHKEDLQREHSOK 146  
DB 187 IYNDERKKNPFRKPKDKNEDKKVPPPSLKEIKNGIDHEENEDKKELMFKLQLLQK 246  
QY 147 SDSTKDVATVLDKNNISSKST 168  
DB 247 QYPLRDIPDPTIRSEYKSMKKT 268

## RESULT 36

E89883  
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: E89883  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaishi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E89883  
A:Status: preliminary  
A:Note: A:Residues: 1-645 <KUR>  
A:Cross-references: UNIPROT:Q99UX5; UNIPARC:UPI00000CAAB3; GB:BA000018; PID:gl13700929; E  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0976

Query Match 10.4%; Score 93.5; DB 2; Length 645;  
Best Local Similarity 21.9%; Pred. No. 27;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
  
QY 11 TTVKGFILNKDTGGE-VSELKPHRVTV--TIQNGKEMSTIVSEEDFILPVYKGELE-KGY 66  
DB 327 SAITEFQNVQPTNEKMDLQDTKYVYVESVENNESMDTFVXH-----PIKTMGLNGKKY 381  
QY 67 QP-----DGWEISGFEGKK-----D 81  
DB 382 MWMTETNDYDKDFWVEGQVRVETISKDAKNVTRTIIPVVEGKTYDALVKVHVTTIDYD 441  
QY 82 AGYVNL-SKDTFIKPVFKKIBEKKEENKPTFDV-----SKKKNQPNVNHSQLNESHK 135  
DB 442 GQYHVRIVKBAFTKANTDKSNKEQDNSAKKEATPATPSKTPSPVEKESOKQDSQKD 501  
QY 136 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISSKSTNNPK 174  
DB 502 DNKLQPSVEKENDASSEGKDKTPATKPTKGEVSESSIT--PTK 543

## RESULT 37

S42488  
dnaK-type molecular chaperone hsp70 - Pyrenomonas salina nucleomorph  
N:Alternate names: heat shock protein 70  
C:Species: nucleomorph Pyrenomonas salina  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C>Date: 13-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 31-Dec-2004  
C:Accession: S42488; S45576  
R:Reising, S.; Hofmann, C.J.B.

submitted to the EMBL Data Library, March 1993  
A:Description: Smallest known eukaryotic genomes encode a protein gene: towards undere  
A:Reference number: S42488  
A:Accession: S42488  
A:Molecule type: DNA  
A:Residues: 1-649 <RES>  
A:Cross-references: UNIPROT:P37899; UNIPARC:UPI000012CC4F; EMBL:X72621; NID:g461335; P  
R:Hofmann, C.J.B.; Reising, S.A.; Haeuber, M.M.; Martin, W.F.; Mueller, S.B.; Couch, J  
Mol. Gen. Genet. 243, 600-604, 1994  
A:Title: The smallest known eukaryotic genomes encode a protein gene: towards an under  
A:Reference number: S45576; MUID:94268506; PMID:8208251  
A:Accession: S45576  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-24,'Q',26-'91','H',93-578 <HOF>  
A:Cross-references: UNIPARC:UPI0000177D1F; EMBL:X72621; NID:g461335  
A:Note: the sequence is revised in GenBank entry PSHSP70, release 111.0, (PIDN:CAA5119  
C:Genetics:  
A:Gene: hsp70  
A:Genome: nucleomorph  
C:Function:  
A:Description: involved in protein folding and assembling/dissassembling of protein com  
C:Superfamily: bcr protein  
C:Keywords: ATP; molecular chaperone; nucleomorph

Query Match 10.4%; Score 93.5; DB 2; Length 649;  
Best Local Similarity 24.5%; Pred. No. 27;  
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;  
  
QY 17 ILNKDTGVSSELKPHRVTVTTIONG---KEMSTIVSEEDFILPVYKGELEKGYQDQWEI 73  
DB 491 ILNVSADKSTGKGNKITITNDKRLSKSEETRMVVEAB-----KYKTEDEK----- 537  
QY 74 SGFEKKDA-----GYVINLS---KDTPIKPVFKKIBEKKEENKPTFDVSKKKNQPNV 125  
DB 538 --LDKLEAKNSLENYAIVNTVRD-----EKLKIQIEDKKSIEEKVK----- 582  
QY 126 HSQLNESHKEDLQREHSQKSDSTKQVTATVLDK 160  
DB 583 --VLEFIETNEDLEKEEYEEKEKELKNPANDPIISK 615

## RESULT 38

G96542  
hypothetical protein F17J6.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:Cross-references: UNIPROT:Q9C6P8; UNIPARC:UPI00000A743C; GB:AB005173; NID:gl11054631;  
C:Genetics:  
A:Gene: F17J6.14  
A:Map position: 1

Query Match 10.4%; Score 93; DB 2; Length 629;  
Best Local Similarity 19.6%; Pred. No. 28;  
Matches 40; Conservative 39; Mismatches 73; Indels 52; Gaps 7;  
  
QY 11 TTVKEFILNKDTGGEV-----SELKPHRVTVTTIONGKEMSTIVSEEDFI 54

Db	391	TTAK---MSSSTAETVQLPAEKGVGRKMDQKVQSQEGPHLETAKPTKDSAMEQIVAEADVAM	447
Qy	55	LPVYGELEKGVQPDGWEISFEGKKGADGVYVNLNSKDTPIKVPVFKIKBEKE-----	106
Db	448	NPIVEKAMSEVBAEGAALNPIVEAEDG-----AMNPIVEKAMSGQIVAEADAALNQAVD	501
Qy	107	---BENKPTFDVSKKKNQ---VNHSQLNESHKEDLQREHSOKSDSKDVTATVLDKN	161
Db	502	ANFQTAQPTGNDADSDDPSEPVSHS---ETLNPPELEKKEVYMRKDATERSVSADCDQKN	558
Qy	162	N-----ISSKSTTNNPNK	174
Db	559	SKLIAESSLQELISASQTENSTQ	582
RESULT 39			
QRMSP1			
Microtubule-associated protein MAP1B - mouse			
N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei			
C/Species: Mus musculus (house mouse)			
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004			
C/Accession: S07549; S44387; A33645			
R/Noble, M.; Lewis, S.A.; Cowan, N.J.			
J. Cell Biol. 109, 3367-3376, 1989			
A/Title: The microtubule binding domain of microtubule-associated protein MAP1B contains			
A/Reference number: A33645; MUID:90094539; PMID:2480963			
A/Accession: S07549			
A/Molecule type: mRNA			
A/Residues: 1-2464 <NOB>			
A/Cross-references: UNIPROT:P14873; UNIPARC:UPI0000297D7; EMBL:X51396; NID:g52999; PIDN			
R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.			
Arch. Biochem. Biophys. 310, 428-432, 1994			
A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.			
A/Reference number: S44387; MUID:94234720; PMID:8179328			
A/Accession: S44387			
A/Status: preliminary			
A/Molecule type: protein			
A/Residues: 653-663, 'IC' <SAN>			
A/Cross-references: UNIPARC:UPI0000173097			
C/Superfamily: microtubule-associated protein MAP1B			
C/Keywords: microtubule binding; phosphoprotein; tandem repeat			
F/589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69			
R-K-E/D-X)			
F/1861-2064/Region: 17-residue repeats			
F/91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph			
F/147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (Co			
F/1953/Binding site: phosphate (Tyr) (covalent) #status predicted			
Query Match 10.4%; Score 93; DB 3; Length 2464;			
Best Local Similarity 27.8%; Pred. No. 1.3e+02;			
Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;			
Qy	14	KEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVVKGELEKGVQPDGWEI	73
Db	584	EKVLVKKDPKVPKTESK-----SVT-----EKEVSS---KEEQ--SPV-KAEVA-----EK	623
Qy	74	SGFEGKKGADGVYVNLNSKDTFIKVPFK-KLEEKKEBENKPTFDVSKKKNQPNVHSQLNES	132
Db	624	QATESKP-----KVTQDKVVKKEIKTKLEEKKE--KPKKEVVKEDKTPL---KKDEK	672
Qy	133	HKEDLQRE-----EHSQKSDSKDQVTVATVLDKNINSSKSTTNNPNK	174
Db	673	PRKEEVKKEIKKEIKKEERKELKGVKKEKTEPLKDAKKEVKKKEKKEKPKK	728
RESULT 40			
S66040			
serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) dacA - Bacillus subtilis			
N/Alternate names: penicillin-binding protein 5			
C/Species: Bacillus subtilis			
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004			

C/Accession: S66040; I39830; A92275; A92307; B61335; D69612; A23307	R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994	
A/Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom	A/Reference number: S65967; MUID:96051385; PMID:7584024
A/Accession: S66040	
A/Status: preliminary; nucleic acid sequence not shown; translation not shown	A/Molecule type: DNA
A/Residues: 1-443 <OG>	
A/Cross-references: UNIPROT:P08750; UNIPARC:UPI000005FDBA; EMBL:D26185; NID:g467326; PID	A/Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993
R/Todd, J.A.; Roberts, A.N.; Johnstone, K.; Piggot, P.J.; Winter, G.; Ellar, D.J.	
J. Bacteriol. 167, 257-264, 1986	A/Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the B
A/Reference number: I39830; MUID:86250602; PMID:3087956	
A/Accession: I39830	A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA	
A/Residues: 48-226 'O', 228-443 <RES>	A/Cross-references: UNIPARC:UPI000016E82C; GB:M13766; NID:g142816; PIDN:AAA22375.1; PID:
R/Waxman, D.J.; Strominger, J.L.	
J. Biol. Chem. 255, 3964-3976, 1980	A/Title: Sequence of active site peptides from the penicillin-sensitive D-alanine carbox
A/Reference number: A92275; MUID:80182289; PMID:6768745	
A/Accession: A92275	A/Molecule type: protein
A/Residues: 32-95, 'X', 97-98, 'XQX', 102 <WAX>	
A/Cross-references: UNIPARC:UPI000017804E	R/Waxman, D.J.; Strominger, J.L.
J. Biol. Chem. 256, 2067-2077, 1981	
A/Title: Primary structure of the COOH-terminal membranous segment of a penicillin-sensi	A/Reference number: A92307; MUID:81117303; PMID:6780559
A/Accession: A92307	
A/Molecule type: protein	A/Residues: 414-443 <WA2>
A/Cross-references: UNIPARC:UPI000017804F	
R/Yocum, R.R.; Waxman, D.J.; Rasmussen, J.R.; Strominger, J.L.	Proc. Natl. Acad. Sci. U.S.A. 76, 2730-2734, 1979
A/Title: Mechanism of penicillin action: penicillin and substrate bind covalently to the	
A/Reference number: A61335; MUID:79223865; PMID:111240	A/Accession: B61335
A/Molecule type: protein	
A/Residues: 55-68 <YOC>	A/Cross-references: UNIPARC:UPI0000178050
R/Kunst, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter	
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho	Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler	
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.P.	Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel	
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, E.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle	Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror	
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,	T.; Winfers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A/Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Banchin, A.	
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.	A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: D69612	
A/Status: nucleic acid sequence not shown; translation not shown	A/Molecule type: DNA
A/Residues: 1-443 <KUN>	
A/Cross-references: UNIPARC:UPI000005FDEA; GB:Z99104; GB:AL009126; NID:g2632267; PIDN:K	A/Experimental source: strain 168
C/Genetics:	
A/Gene: dacA	A/Start codon: TTG
C/Superfamily: penicillin-binding protein 5	
C/Keywords: cell wall synthesis; hydrolase; serine carboxypeptidase	F/67/Active site: Ser #status experimental





**THIS PAGE BLANK (uspio)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 105.685 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KLVKDFARTTIVKSFILNK.....ATVLDKNISSKSTNNPNK 174

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	897	100.0	2119	Q9AHT5_STRPN	Q9Aht5 streptococc
2	897	100.0	2140	Q97RY6_STRPN	Q97ry6 streptococc
3	894	99.7	2144	Q8DQP7_STRR6	Q8dqp7 streptococc
4	893	99.6	2144	Q9S4M8_STRPN	Q9s4m8 streptococc
5	139	15.5	300	Q4XUI6_PLACH	Q4xui6 plasmodium
6	119	13.3	775	Q8CFK8_STABP	Q8cfk8 staphylococ
7	117	13.0	1069	Q512T7_ENTHI	Q512t7 entamoeba h
8	115.5	12.9	296	Q50LX8_ENTHI	Q50lx8 entamoeba h
9	114.5	12.8	1038	Q90784_CHICK	Q90784 gallus gall
10	114	12.7	609	Q812K8_PLAF7	Q812k8 plasmodium
11	112.5	12.5	346	Q9U0G0_PLARE	Q9u0g0 plasmodium
12	112.5	12.5	354	Q2S995_PLAFA	Q2s995 plasmodium
13	112.5	12.5	354	Q81J55_PLAF7	Q81j55 plasmodium
14	112.5	12.5	361	Q9SP15_PLAFA	Q9sp15 plasmodium
15	112.5	12.5	379	Q9U6C4_PLAFA	Q9u6c4 plasmodium
16	112.5	12.5	662	Q4YMU4_PLABE	Q4ymu4 plasmodium
17	112	12.5	829	Q815P3_PLAF7	Q815p3 plasmodium
18	111.5	12.4	379	Q2S706_PLAFA	Q2s706 plasmodium
19	111.5	12.4	775	Q5HQ11_STABQ	Q5hq11 staphylococ
20	110.5	12.3	379	Q2S705_PLAFA	Q2s705 plasmodium
21	110	12.3	380	Q2G019_PLAFA	Q2g019 plasmodium
22	110	12.3	500	Q6BGL7_PARTB	Q6bgl7 paramacium
23	110	12.3	674	Q7RLE7_PLAYO	Q7rle7 plasmodium
24	109.5	12.2	384	Q50VJ0_ENTHI	Q50vj0 entamoeba h
25	109.5	12.2	3008	Q81436_PLAF7	Q81436 plasmodium
26	109	12.2	951	Q96229_PLAF7	Q96229 plasmodium
27	108.5	12.1	600	Q77355_PLAF7	Q77355 plasmodium
28	108.5	12.1	973	Q7RB37_PLAYO	Q7rb37 plasmodium
29	108	12.0	467	Q59PE2_CANAL	Q59pe2 candida alb
30	108	12.0	467	Q59PL2_CANAL	Q59pl2 candida alb
31	107.5	12.0	470	Q9FJK9_ARATH	Q9fjk9 arabidopsis

32	107.5	12.0	948	2	Q9U4U6_PLAFA	Q9u4u6 plasmodium
33	107.5	12.0	1455	2	Q640L5_MOUSE	Q640l5 mus musculu
34	106.5	11.9	374	2	Q5V9M0_PLAKN	Q5v9m0 plasmodium
35	106.5	11.9	1015	2	Q5W5T1_TETPY	Q5w5t1 tetrahymena
36	106	11.8	616	2	Q6BRW2_DBBHA	Q6brw2 debaromyce
37	106	11.8	988	2	Q7RP53_PLAYO	Q7rp53 plasmodium
38	106	11.8	1268	2	Q54HA7_DICDI	Q54ha7 dictyosteli
39	106	11.8	3026	2	Q8ILS9_PLAF7	Q8ils9 plasmodium
40	105.5	11.8	329	2	Q9NFV9_PLAFA	Q9nf9 plasmodium
41	105.5	11.8	605	2	Q7RJ33_PLAYO	Q7rj33 plasmodium
42	105.5	11.8	736	2	Q4VYV2_PLABE	Q4vyv2 plasmodium
43	105.5	11.8	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
44	105	11.7	1550	2	Q54GSI_DICDI	Q54gsi dictyosteli
45	105	11.7	3127	2	Q8IDA0_PLAF7	Q8ida0 plasmodium

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=11116976; PubMed=11179332;  
DOI=10.1128/JAI.69.3.1593-1598.2001;  
RA Wozniak T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S.  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection."  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0003086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Protease.

FT NON TER 1 1  
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 897; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 1.5e-52;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPEILNKDTGVSSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKG 60  
Db 1912 KIVVKDFARNTTVKPEILNKDTGVSSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKG 1971  
QY 61 ELEKGYPFGWISGPEGKDDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKDD 120  
Db 1972 ELEKGYPFGWISGPEGKDDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKDD 2031  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
Db 2032 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2085

RESULT 2  
Q97RY6 STRPN  
ID Q97RY6 STRPN PRELIMINARY; PRT; 2140 AA.  
AC Q97RY6;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Serine protease, subtilase family.  
GN OrderedLocusNames=SP0641;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC BAA-334 / TIGR4;  
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M.B., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,  
RA Hickey E.K., Holt I.E., Lofus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL; AS007373; AAK74791.1; -; Genomic\_DNA.  
DR PIR; F95074; F95074.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR TIGR; SP0641; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR002029; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.

TIGRFams; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Complete proteome; Protease.  
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 897; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 1.5e-52;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPEILNKDTGVSSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKG 60  
Db 1933 KIVVKDFARNTTVKPEILNKDTGVSSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKG 1992  
QY 61 ELEKGYPFGWISGPEGKDDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKDD 120  
Db 1993 ELEKGYPFGWISGPEGKDDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKDD 2052  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
Db 2053 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 3  
Q8DQP7 STRR6  
ID Q8DQP7 STRR6 PRELIMINARY; PRT; 2144 AA.  
AC Q8DQP7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).  
GN Name=prtA; OrderedLocusNames=spr0561;  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21429245; PubMed=11544234;  
DOI=10.1128/JB.183.19.5709-5717.2001;  
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geisinger C.,  
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushina P.,  
RA McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,  
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
RA Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AE008434; AAK9365.1; -; Genomic\_DNA.  
DR PIR; A97942; A97942.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR002029; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.

[illegible]

```

RESULT 5
Q4XU16_PLACH PRELIMINARY; PRT; 300 AA.
ID Q4XU16 PLACH PRELIMINARY;
AC Q4XU16;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PC000286.03;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCB1_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N.L., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriaman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the plasmodium life cycle by genomic,
RL transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC ENBL; CAJ01003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
NN 1
FT NON TER
SQ SEQUENCE 300 AA; 34469 MW; 8ED4E512AFB1945 CRC64;
Query Match 15.5%; Score 139; DB 2; Length 300;
Best Local Similarity 23.2%; Pred.No. 0.11;
Matches 45; Conservative 38; Mismatches 61; Indels 50; Gaps 8;
QY 1 KIYVDDFARNITVKFIFLNKDTGE-----VSELKPHRVTVTIQNGKEM 43

```

```
Db      85 KTIYNNFLKNQNHKLVSGDDGTLHLIELCKNDKYKSTRLLGHQKPVHTQFSNGKFI 144
QY      44 STIYSE-----EDFILPVYKGELEKGYQFDGWEISGPEKKDAGVIVNLSKDTIK 95
Db      145 ASSSFDKSTRIWIGDITGYLAVYRGHVGPAYKI-AWSI-----DNNYIYSCSDSTLK 196
QY      96 -----PVFKKIEKKE---EENKPTFVDSKKDNPVNHSQLNESHKEDLQREHS 144
Db      197 LWRINHLVLLKXKENAEQTDEQK-----NEQKENTQ-NNDQPNDEANSEKKKKNEK 250
QY      145 QKSDSTKVATVL 158
Db      251 ERNDKTKNKIKTL 264

RESULT 6
Q8CPK8 STABP
ID Q8CPK8 STABP PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocusNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1284;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AE016746; RAO04453.1; -; Genomic_DNA.
DR HSSP; FI4677; IQME.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensus Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP dimer.
DR InterPro; IPR012338; PBP_trypt_fold.
DR InterPro; IPR001460; Pencil_bind_trypt.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 13.3%; Score 119; DB 2; Length 775;
Best Local Similarity 27.0%; Pred. No. 7.2;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 20 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GSEKGYQFDGW-- 71
Db 614 EDSVNAQSLKP---ITIGNRQIKQSVKSKVLPKPHSKVLMMLTDGLTNP-DMTGWTK 668
QY 72 -HISGPE-----GKIDAGYIN--LSKDTIKPVFKKIEKKEENKPTFVDS----K 117
Db 669 EDVLAFEDLTIKVSTKNGFVNTQISKGQIK-----NKKIEVLSLSD 715
QY 118 KKDNPVNHSQLNESHKEDLQREHSQKSDSTKVATVLKDNKISSKSTNN 171
Db 716 TDDQDEKTDSDNKSCKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 763

RESULT 7
Q512T7 ENTHI
ID Q512T7 ENTHI PRELIMINARY; PRT; 1069 AA.
```

```
AC Q512T7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Amedeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Church C., Hance Z., Harris B., Whitehead S.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Wang Z.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Whitehead S.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichertitz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;

Query Match 13.0%; Score 117; DB 2; Length 1069;
Best Local Similarity 25.6%; Pred. No. 14;
Matches 50; Conservative 39; Mismatches 70; Indels 36; Gaps 9;

QY 1 KIVVKDPARNTT-----VKEFILNKDT-GEVSELKPHRVTVTIQNG-KEMSTIVSEED 52
Db 108 EIEQKEIEKNTVSTIEDSVKSLDSDSKQSLPEVSKDNVNESSTLTNGEKKLSLCLNEQD 167
QY 53 FILPVYKGELEKGYQFDGWEISGPEKKDAGVIVNLSKDTIKPVFKKIEKKEENK-- 110
Db 168 -----ELQSKSSSTD--NKNKRDRIHFVDVLPKNEKEISMEIESKTEEEKSN 217
QY 111 ---PTFDVSKKKKNPQ-----VNHSQLNESHKEDLQREHS-QKSDSTKD-----VT 154
Db 218 LQIFSLNLSEKGRKNVSEIATVILKSNSSNGEEDKQDEEVCERKPFDSQEEKKEWIK 277
QY 155 ATVLDDKNNISKSTT 169
Db 278 AEVSQNKVEKDKSTT 292

RESULT 8
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Amedeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
```



```

DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON_TER 1 346
FT SEQUENCE 346 AA; 39127 MW; A804B96BDPAFA010 CRC64;
SQ
Query Match 12.5%; Score 112.5; DB 2; Length 346;
Best Local Similarity 23.7%; Pred. No. 8.4;
Matches 42; Conservative 33; Mismatches 67; Indels 35; Gaps 7;

QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQF 68
DB 156 STKTKEYAQKNAVEKAKNAQKAVLKAKKASS-----YNYL----- 197

QY 69 DGWEISGF--EGKKDAG-----YVINLSKDTFIKVPFKKIEEKKEE-----E 108
DB 195 -GWFEFGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEELEEKNE 253

QY 109 NKPTFDVSKKDNFQVNHSQLNESHKEDLQREHBSQKSDTKDVTATVLDKNNISSKST 168
DB 254 EETSEISEDEBEKEEENDEKKEQKQESQNNNDQKQKQMEAE-----QNLISKQ 308

QY 169 TNN 171
DB 309 NNN 311

RESULT 13
Q81J55_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN ORFNames=PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12369864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S., Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum".
RL Nature 415:498-511(2002).
DR EMBL; AB014834; AAN35542.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 12.5%; Score 112.5; DB 2; Length 354;
Best Local Similarity 22.4%; Pred. No. 8.6;
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQF 68
DB 153 STKTKEYAQKNAVEKAKNAQKAVLKAKKASS-----YNYL----- 194

QY 69 DGWEISGF--EGKKDAG-----YVINLSKDTFIKVPFKKIEEKKEE-----E 108
DB 195 -GWFEFGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEELEEKNE 253

QY 109 NKPTFDVSKKDNFQVNHSQLNESHKEDLQREHBSQKSDTKDVTATVLDKNNISSKST 168
DB 254 EETSEISEDEBEKEEENDEKKEQKQESQNNNDQKQKQMEAE-----QNLISKQ 308

QY 169 TNN 171
DB 309 NNN 311

RESULT 14

```



Query N  
Best Lc

608 K-----KKKKERNDKIKSKIKTLL 626

Db

RESULT 17

Q815F3\_P1AF7

ID Q815F3\_P1AF7 PRELIMINARY; PRT; 829 AA.

AC Q815F3;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN ORFNames=PFL1275C;

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=36329;

LN [1]

RP NUCLEOTIDE SEQUENCE.

EX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S., Portea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrall B.G.;

RA "Genome sequence of the human malaria parasite Plasmodium falciparum.";

RT Nature 419:498-511(2002).

RL Nature 419:498-511(2002).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M., Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;

RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.

RL EMBL; AE014848; AAN36341.1; -; Genomic\_DNA.

KW Hypothetical protein.

SQ SEQUENCE 829 AA; 98815 MW; EF2675E301B2CE93 CRC64;

Query Match 12.5%; Score 112; DB 2; Length 829;

Best Local Similarity 24.0%; Pred. No. 23;

Matches 44; Conservative 39; Mismatches 58; Indels 42; Gaps 9;

QY 8 ARNTTKVETLNKDTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVVKGBLEKGYQ 67

Db 487 AATITSDNSILNKNKKI--IQP--LEYLLKNIGD--KTLTERDIVLDVFPYMKYKH 540

QY 68 FDGWEISGPEGKADGYVNLKDTFKPKVKKIEKKEENKPT-----F 113

Db 541 LN-----KKETLNFSLN-----PREIEKNRDKKKGTHNNKNDASEYMLKY 583

QY 114 DVSKKXON-PQVNHSSOLNESH-RKEDLQREHSQKSDSTKDTATVLDKNNISKSTNN 171

Db 584 KIKKKKKNFENNTELNSNKKNNKLVER---DNSLKQEQIILINDKNVIEHTKIYDN 640

QY 172 PNK 174

Db 641 QKK 643

RESULT 18

ID Q25706\_PLAFA

AC Q25706\_PLAFA PRELIMINARY; PRT; 379 AA.

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Polymorphic antigen.

GN Plasmodium falciparum.

OS Plasmodium.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5833;

LN [1]

RP NUCLEOTIDE SEQUENCE.

QY 20 KDTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVYK-----GELEKGYQDGV-- 71

Db

RESULT 19

Q5HQ11\_STAEO

ID Q5HQ11\_STAEO PRELIMINARY; PRT; 775 AA.

AC Q5HQ11;

DT 10-MAY-2005 (TrEMBLrel. 30, Created)

DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Penicillin-binding protein 1.

GN Name=pbpi; OrderedLocusNames=SERP0746;

OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=176279;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RA PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;

RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T., Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J., Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S., Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C., Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H., Hance I.R., Nelson K.E., Fraser C.M.;

RA "Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain.";

RT J. Bacteriol. 187:2426-2438(2005).

RL J. Bacteriol. 187:2426-2438(2005).

DR EMBL; CP000029; AAW54126.1; -; Genomic\_DNA.

DR TIGR; SERP0746; -.

DR GO; GO:0008658; F:penicillin binding; IEA.

DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.

DR InterPro; IPR005543; PASTA.

DR InterPro; IPR005311; PBP dimer.

DR InterPro; IPR001460; Pencil\_bind\_tpept.

DR Pfam; PF03793; PASTA; 2.

DR Pfam; PF03717; PBP dimer; 1.

DR Pfam; PF0905; Transpeptidase; 1.

DR SMART; SM00740; PASTA; 2.

DR Complete proteome.

KW Complete proteome.

SQ SEQUENCE 775 AA; 86352 MW; B9395893B0043694 CRC64;

Query Match 12.4%; Score 111.5; DB 2; Length 775;

Best Local Similarity 24.9%; Pred. No. 23;

Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 20 KDTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVYK-----GELEKGYQDGV-- 71

Db

RESULT 20

Q5HQ11\_STAEO

ID Q5HQ11\_STAEO PRELIMINARY; PRT; 775 AA.

AC Q5HQ11;

DT 10-MAY-2005 (TrEMBLrel. 30, Created)

DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Penicillin-binding protein 1.

GN Name=pbpi; OrderedLocusNames=SERP0746;

OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=176279;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RA PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;

RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T., Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J., Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S., Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C., Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H., Hance I.R., Nelson K.E., Fraser C.M.;

RA "Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain.";

RT J. Bacteriol. 187:2426-2438(2005).

RL J. Bacteriol. 187:2426-2438(2005).

DR EMBL; CP000029; AAW54126.1; -; Genomic\_DNA.

DR TIGR; SERP0746; -.

DR GO; GO:0008658; F:penicillin binding; IEA.

DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.

DR InterPro; IPR005543; PASTA.

DR InterPro; IPR005311; PBP dimer.

DR InterPro; IPR001460; Pencil\_bind\_tpept.

DR Pfam; PF03793; PASTA; 2.

DR Pfam; PF03717; PBP dimer; 1.

DR Pfam; PF0905; Transpeptidase; 1.

DR SMART; SM00740; PASTA; 2.

DR Complete proteome.

KW Complete proteome.

SQ SEQUENCE 775 AA; 86352 MW; B9395893B0043694 CRC64;

Query Match 12.4%; Score 111.5; DB 2; Length 775;

Best Local Similarity 24.9%; Pred. No. 23;

Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 20 KDTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVYK-----GELEKGYQDGV-- 71

Db

RESULT 21

Q5HQ11\_STAEO

ID Q5HQ11\_STAEO PRELIMINARY; PRT; 775 AA.

AC Q5HQ11;

DT 10-MAY-2005 (TrEMBLrel. 30, Created)

DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Penicillin-binding protein 1.

GN Name=pbpi; OrderedLocusNames=SERP0746;

OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=176279;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RA PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;

RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T., Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J., Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S., Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C., Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H., Hance I.R., Nelson K.E., Fraser C.M.;

RA "Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain.";

RT J. Bacteriol. 187:2426-

```
Db 614 EDSVNAQSLKP-----ITIGNKQIKQSVKSGTKVLPHPKSVMLMTDGLTMP-DMTGWTX 668
QY 72 -BISGFE-----GKDRAGYVIN--LSKDTPIKPVFK-----KIEKKEBENKPTF 113
Db 669 EDVLAPEDLTKIKVSTKNGFGVTNQSISKGQIKKNDKIEVLSAEDTDDQEKTDSS 728
QY 114 DYSKKKONQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNN 162
Db 729 DKSKKDKVDEDSNASSSKNEKSNADSKNDSDDSTNETSGS--ERN 775

RESULT 20
Q25705 PLAPA
ID Q25705 PLAPA PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08851; AAC47831.1; -; Unassigned_DNA.
DR InterPro; IPRO10784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C9AA0 CRC64;

Query Match 12.3%; Score 110.5; DB 2; Length 379;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;

QY 21 DTGGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGY----- 67
Db 167 ETGE-----RNSRNFVTTKTB-----YAGKVKDYERAKNAYQKANQAV 207
QY 68 -----PD-----GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEBEN 109
Db 208 LKAKEASSDYTLGWEFGGVPHEKKENMLSHLYSSKDKENISKENDVDLDE-KBEEA 266
QY 110 KPTDPSVKKKONPVNHSQNLN-----ESHRKEDIQREHSQKSDSTKDVAT 156
Db 267 ETEEBEEKKEETESISEDEDEEBEKEBKEKKEKQEQEKQSQNNNDQKKDMEA- 325

QY 157 VLDKNNISKSTNN 171
Db 326 ----QNLISKNNNN 336

RESULT 21
Q26019 PLAPA
ID Q26019 PLAPA PRELIMINARY; PRT; 380 AA.
AC Q26019;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.D., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
```

```
RT Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L07944; AAC09378.1; -; Genomic_DNA.
DR PDB; 1PSM; NMR; @=90-127.
DR InterPro; IPRO10784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
DR Signal.
KW SIGNAL.
FT CHAIN 26 380 polymorphic antigen.
SQ SEQUENCE 380 AA; 43290 MW; 0986CA1393094CA2 CRC64;

Query Match 12.3%; Score 110; DB 2; Length 380;
Best Local Similarity 24.2%; Pred. No. 14;
Matches 45; Conservative 34; Mismatches 59; Indels 48; Gaps 9;

QY 7 FARNTTVKKEFILN-KDTGGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEK 65
Db 179 FSRKT--KEYAEKAKNAYEKANAYQKANQAVLKAKKASS-----YDYL----- 221
QY 66 YQFQGWETISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEBENKPTFDVSKK 118
Db 222 ----GWEPGGVPEHKKENMLSHLYSSKDKENISKENDVDLDE-KBEEAEETEEL 276
QY 119 KDNQVNHSQLN-----ESHRKEDIQREHSQKSDSTKDVATVLDKNNISS 165
Db 277 EKNBEETESISEDEDEEBEKEBKEKKEKQEQEKQSQNNNDQKKDMEA-----QNLIS 331

QY 166 KSTNN 171
Db 332 KNQNNN 337

RESULT 22
Q68GL7 PARTE
ID Q68GL7 PARTE PRELIMINARY; PRT; 500 AA.
AC Q68GL7;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PTMB.06c;
OS Paramesium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramesium.
OX NCBI_TaxID=5888;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Stock d4-2;
RX PubMed=15296759; DOI=10.1016/j.cub.2004.07.029;
RA Zagulski M., Nowak J.K., Le Mouel A., Nowacki M., Migdalski A.,
Gromadka R., Noel B., Blanc I., Dessein P., Wincker P., Keller A.M.,
Cohen J., Meyer E., Sperling L.;
RT "High Coding Density on the Largest Paramesium tetraurelia Somatic
Chromosome.";
RL Curr. Biol. 14:1397-1404(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Stock d4-2;
RA Nowak J.K., Migdalski A., Gromadka R., Zagulski M.;
RT "Paramesium megabase sequencing project.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR548612; CAH03203.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 500 AA; 56364 MW; DB4D7F90C86E79F0 CRC64;
```

```
Query Match          12.3%; Score 110; DB 2; Length 500;
Best Local Similarity 22.4%; Pred. NO. 19;
Matches      43; Conservative 34; Mismatches       75; Indels   40; Gaps    6;
```

/

D    /         14 KEPIFKDTGTVSELKPHRVVTVTTQNGKESS-----TIVSEDFILPVYKG 60  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
D    D      11 KODLLKKGGAKSIRE-BFFQTVP IQNAEESEKCKLKVWNEHMTATSDDKILLIDATLG 69  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
/            61 ELEKYGFQDGWEISFGFEKGDAGVVIN----LSKDTPFIKPVEK-----KIEEKK 105  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
D          70 NEEAQREVIQIYTFANVQRSIAITQBGVQPVRPDDFVENFSPQMOKINKUKIEROR 129  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
/          106 EENKNPTFDVSX---KKONPQNHSQLNESHKR-----EDLRBEHSQSXSSTXDVT 154  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
D          130 DEKNKKEEKMKGHSGSKLKWKMIKKMEHEKHQRNKVALEQWGGEIKEGSKDKARDLD 189  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
/          155 ATVLDNKNISSK 166  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
  
D          190 EIHKNNQISKK 201  
| : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 23

QRLRLET PLAYO PRELIMINARY; PRT; 674 AA.

D Q7RLR7;  
C Q7RLR7;  
A C Q7RLR7;  
A A Silva J.C., Emolaeva S.V., Suh B.B., Koelij T.W., Pertea M., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V., Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M., Florens L., Yates J.R. III, Rainey J.D., Sindén R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., van Lin H.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;  
C "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";  
C Nature 419:512-519(2002).  
C -! CAUTION: The sequence shown here is derived from an preliminary data.  
C EMBL; AABL01000712; EAA22065.1; -, Genomic\_DNA.  
C HSSP; P16649; 1ERJ.  
C InterPro; IPR001680; WD40.  
C Pfam; PF00400; WD40; 8.  
C PRINTS; PR003320; GPSTEINSBPT.  
C ProDom; PD000018; WD40; 4.  
C PROSITE; PS00678; WD REPEATS\_1; 2.  
C PROSITE; PS00082; WD REPEATS\_2; 6.  
C PROSITE; PS02994; WD\_REPEATS\_REGION; 2.  
C Repeat; WD repeat.  
C SEQUENCE 674 AA; 76535 MW; E7521B469FE80F7 CRC64;

```

Qy  44 SSTIVSE-----EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIK 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  505 ASSSFDKSIIRWSGIDGTYLAVPRGHVPAYKI-AWSI-----DNNYIISCSQDSTLK 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  96 -----PVPKIEE-----KKEENKPTFDV--SKKDNQPVNHSQLNESHKRE 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  557 LWRINHLVPLTKKKEBNGDQPSDPKNDQKDDXKDDQKQKNDNNDHANNQEBGE 616
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  137 DLQREHSQKSDSTKQDVATATVL 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  617 EKKKKKEKKADKIKSIKTLL 638
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 24
Q50VJ0_ENTHI PRELIMINARY; PRT; 384 AA.
AC Q50VJ0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE HMG box protein.
GN ORFNames=188.t00012;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC SPRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jegels K., Moule S., Mungall K., Omond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foister P.G., Sichertitz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RA "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF01000585; EAL45607.1; -; Genomic DNA.
SQ SEQUENCE 384 AA; 45464 MW; 207789F65D72E019 CRC64;

Query Match 12.2%; Score 109.5; DB 2; Length 384;
Best Local Similarity 27.1%; Pred. No. 15;
Matches 39; Conservative 28; Mismatches 64; Indels 13; Gaps 4

Qy  27 ELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFD--GWEISGPEGKQDAGY 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  226 EERTKVKVEIKKDDKTKVKEIKED-----EKKEKHKSKKCKKEEMKKNKGEKESDK 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  85 VINLSKDTFIKVPVKTKIEEKEENKPTFDVSKKDNQPVNHSQLNESHKREDLQREHS 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  281 KEDTKDKKKVKKSEKQDQIKKEDSKH-----EKKEETEEKKPKPKPEKESKKEKKHS 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  145 QKSDSTKD--VTATVLDKNINSSK 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  337 KKEDKKDEKSKKVEDKSKKQK 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 25
Q8I436_PLAF7
ID Q8I436_PLAF7 PRELIMINARY; PRT; 3008 AA.
AC Q8I436;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PFE0325W.

```



















**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 102.913 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TVVKEFILNKDTGVSELKP.....ATVLDKNMISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*
- 9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	4	AAB48343 S. pneumo
2	848	100.0	2120	3	AAY81710 Streptoco
3	848	100.0	2140	6	ABU01020 S. pneumo
4	848	100.0	2140	6	ABU45746 Protein e
5	848	100.0	2140	8	ADM92113 S. pneumon
6	848	100.0	2140	8	ADM50099 S. pneumon
7	845	99.6	637	8	Adt-94534 Novel S.
8	845	99.6	637	9	Aes58404 Streptoco
9	845	99.6	2138	8	Adk48759 Streptoco
10	615	72.5	117	2	AAM55096 Streptoco
11	615	72.5	117	5	ABP54590 S. pneumo
12	615	72.5	117	7	Adc45149 S. pneumo
13	119	14.0	746	5	AAG81779 S. epider
14	119	14.0	778	5	ABP39023 Staphyloc
15	119	14.0	778	8	AdS06368 Staphyloc
16	111.5	13.1	354	9	Adz72253 Plasmodiu
17	111.5	13.1	775	6	ABU42797 Protein e
18	111	13.1	707	6	ABU25018 Protein e
19	110.5	13.0	647	9	Adz79635 P. falcip
20	110.5	13.0	651	8	Ado19012 Amino aci
21	110.5	13.0	651	8	Ado19010 P. falcip
22	108	12.7	188	9	Adz79639 P. falcip
23	107.5	12.7	470	8	Adt56185 Plant pol
24	107.5	12.7	484	3	AAG47777 Arabidops

25	106	12.5	665	3	AAB18278	Plasmodiu
26	106	12.5	665	7	ABO23606	Plasmodiu
27	103.5	12.2	169	9	ADZ79634	P. falcip
28	103	12.1	903	6	ABU24404	Protein e
29	101.5	12.0	564	4	ABB61977	Drosophil
30	100	11.8	1791	8	ADP25441	Plasmodiu
31	99.5	11.7	1384	6	ABP55413	Human MDD
32	99.5	11.7	1404	6	ABP55393	Human MDD
33	97.5	11.5	645	9	ADW88460	Staphyloc
34	97.5	11.5	645	9	ADW88459	Staphyloc
35	97.5	11.5	645	9	ADW88458	Staphyloc
36	97.5	11.5	1184	6	ABU25330	Protein e
37	97.5	11.5	2468	6	ABR64281	Anglogene
38	97.5	11.5	2468	7	ADE62723	Human Pro
39	97.5	11.5	2468	7	ADE62719	Human Pro
40	97.5	11.5	2468	7	ADE62727	Human Pro
41	97.5	11.5	2468	7	ADE62715	Human Pro
42	97.5	11.5	2468	8	ADL12997	Human ste
43	97.5	11.5	2468	8	ADN05260	Antipsori
44	97.5	11.5	2468	8	ADR14614	Human NP-
45	97.5	11.5	2519	4	ABG16636	Novel hum

#### ALIGNMENTS

##### RESULT 1

AAB48343

ID AAB48343 standard; protein; 773 AA.

XX AAB48343;

AC AAB48343;

DT 20-APR-2001 (first entry)

XX S. pneumoniae Spi30 polypeptide.

XX Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

XX WO200076540-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US015925.

XX 10-JUN-1999; 99US-0138453P.

XX (MEDI-) MED IMMUNE INC.

XX Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

XX N-PSDB; AAC84742.

XX New vaccines comprising Spi28 or Spi30 polypeptides, for treating and  
preventing pneumococcal infections, particularly infections caused by  
Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
blood infections.

Claim 8; Page 51-54; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
the treatment and prevention of pneumococcal infections, particularly  
infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
bronchial, lung or blood infections. The antigens are used as immunogenic  
agents to stimulate an immune response. The antisera and antibodies may  
also be used in diagnosing and treating pneumococcal infections.  
XX Recombinant polypeptides serve as a mechanism for stimulating production  
of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Sp130 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 848; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 2.8e-74;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKGFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db 610 TTVKGFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 669  
QY 61 WISGFGKKGADAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 120  
Db 670 WISGFGKKGADAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 729  
QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 164  
Db 730 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW Bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
XX 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
XX (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
XX Le Page RWP, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
XX WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
XX Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
XX Claim 2; Page 41-42; 76pp; English.  
XX  
XX This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 848; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 1.1e-73;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKGFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db 1923 TTVKGFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 1982  
QY 61 WISGFGKKGADAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 120  
Db 1983 WISGFGKKGADAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 2042  
QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 164  
Db 2043 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
XX antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
FN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Masignani V, Tettelin H, Fraser C;  
XX  
XX WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
XX Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target



CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)

XX Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVAKEFILNKDTGEVSELKPHRVTTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1943 TTVAKEFILNKDTGEVSELKPHRVTTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKKDAGYVNLKSDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 120  
 DB 2003 WEISGFEKKDAGYVNLKSDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 164  
 DB 2063 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.

XX ABU45746;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #31273.  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX Streptococcus pneumoniae.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 XX 25-OCT-2001; 2001US-0342923P.  
 XX 08-FEB-2002; 2002US-00072851.  
 XX 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 73670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVAKEFILNKDTGEVSELKPHRVTTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1943 TTVAKEFILNKDTGEVSELKPHRVTTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKKDAGYVNLKSDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 120  
 DB 2003 WEISGFEKKDAGYVNLKSDTFIKPVFKIIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 164  
 DB 2063 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.

XX ADM92113;  
 XX 03-JUN-2004 (first entry)  
 XX S pneumoniae antigenic protein sequence SeqID310.  
 XX antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 XX antigenic.  
 XX Streptococcus pneumoniae.

PN WO2004020609-A2.  
XX  
XX PD 11-MAR-2004.  
XX PF 02-SEP-2003; 2003WO-US027401.  
XX PR 30-AUG-2002; 2002US-0407082P.  
XX PA (TUFT ) UNIV TUFTS.  
XX PI Camilli A, Hava DL;  
XX WPI; 2004-239189/22.  
XX DR N-PSDB; ADM91876.  
XX  
XX New Streptococcus pneumoniae nucleic acid molecules, useful for  
PT diagnosing, treating and preventing active infections of Streptococcus  
PT pneumoniae.  
XX  
XX Claim 27; SEQ ID NO 310; 123pp; English.  
XX  
XX This invention relates to novel isolated Streptococcus pneumoniae nucleic  
CC acid molecules and the antigenic polypeptides encoded by them. The  
CC invention may be useful for the production of compounds with an  
CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
CC compositions and methods disclosed are useful for treating Streptococcus  
CC pneumoniae infection. The present sequence is that of an S pneumoniae  
CC protein of the invention.  
XX  
XX Sequence 2140 AA;  
SQ  
Query Match 100.0%; Score 848; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 1943 TTVEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002  
QY 61 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKNPQVNHSQLN 120  
DB 2003 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKNPQVNHSQLN 2062  
QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 164  
DB 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 2106  
RESULT 6  
ADP50099  
ID ADT50099 standard; protein; 2140 AA.  
XX AC ADT50099;  
XX  
XX 13-JAN-2005 (first entry)  
XX  
DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
XX  
KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
KW sepsis; meningitis.  
XX  
OS Streptococcus pneumoniae TIGR4.  
XX  
XX WO2004092209-A2.  
XX  
XX PD 28-OCT-2004.  
XX  
XX 15-APR-2004; 2004WO-EP003984.  
XX PF  
XX 15-APR-2003; 2003EP-00450087.  
XX PR  
XX (INTE-) INTERCELL AG.  
PA

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
PI  
XX WPI; 2004-758335/74.  
XX DR N-PSDB; ADT49955.  
XX  
XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
PT treating S. pneumoniae infections.  
XX  
XX Disclosure; SEQ ID NO 177; 191pp; English.  
XX  
XX This invention relates to novel nucleic acids encoding hyperimmune serum  
CC reactive antigens, or fragments derived thereof. Specifically, it refers  
CC to antigens selected from peptides and serum reactive epitopes that can  
CC be used in pharmaceutical compositions that exhibit antibacterial  
CC activity. The present invention describes a composition (including the  
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
CC that is useful for manufacturing a medicament such as a vaccine, which  
CC can be used to treat or prevent bacterial infections, particularly S.  
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
CC be used for isolating, purifying and/or identifying an interaction  
CC partner of the hyperimmune serum reactive antigen, as well as for  
CC manufacturing a functional nucleic acid selected from aptamers and  
CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
CC of the invention.  
XX  
XX Sequence 2140 AA;  
SQ  
Query Match 100.0%; Score 848; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 1943 TTVEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002  
QY 61 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKNPQVNHSQLN 120  
DB 2003 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKNPQVNHSQLN 2062  
QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 164  
DB 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 2106  
RESULT 7  
ADR94534  
ID ADR94534 standard; protein; 637 AA.  
XX AC ADR94534;  
XX  
XX 16-DEC-2004 (first entry)  
XX  
XX Novel S. pneumoniae protein sequence, SEQ ID 3169.  
XX  
XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
KW bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX US6800744-B1.  
XX PN  
XX PD 05-OCT-2004.  
XX  
XX 30-JUN-1998; 98US-00107433.  
XX PF  
XX 02-JUL-1997; 97US-0051553P.  
XX PR 12-MAY-1998; 98US-0085131P.  
XX

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX Doucette-Stamm LA, Bush D;  
XX WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
XX  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
XX Sequence 637 AA;  
Query Match 99.6%; Score 845; DB 8; Length 637;  
Best Local Similarity 99.4%; Pred. No. 4.3e-74;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVKEFILNKDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
Db |||||  
QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKONPQVNHSQLN 120  
Db |||||  
QY 500 WEISGFEKGDAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKONPQVNHSQLN 559  
QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 164  
Db |||||  
QY 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 603  
Db |||||  
RESULT 8  
ID AEA58404 standard; protein; 637 AA.  
XX AEA58404;  
AC AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
DT  
DE Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
XX  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
KW vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX US2005136404-A1.  
PN  
XX

PD 23-JUN-2005.  
XX  
XX 10-JUL-2003; 2003US-00617320.  
XX  
XX 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
XX (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
XX New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
XX Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
XX Sequence 637 AA;  
Query Match 99.6%; Score 845; DB 9; Length 637;  
Best Local Similarity 99.4%; Pred. No. 4.3e-74;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVKEFILNKDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
Db |||||  
QY 440 TTVKEFILNKDTGEVSELKPHRVTTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 499  
Db |||||  
QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKONPQVNHSQLN 120  
Db |||||  
QY 500 WEISGFEKGDAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSKKONPQVNHSQLN 559  
QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 164  
Db |||||  
QY 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 603  
Db |||||  
RESULT 9  
ID ADK48759 standard; protein; 2138 AA.  
PN  
XX

```
XX ADK48759;
XX
XX 20-MAY-2004 (first entry)
XX
XX Streptococcus pneumoniae protein, Seq ID No 5274.
XX
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
XX Streptococcus pneumoniae.
XX
XX US6699703-B1.
XX
XX 02-MAR-2004.
XX
XX 26-MAY-2000; 2000US-00593110.
XX
XX 02-JUL-1997; 97US-0051553P.
XX
XX 12-MAY-1998; 98US-0085131P.
XX
XX 30-JUN-1998; 98US-00107433.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
XX WPI; 2004-212399/20.
XX
XX N-PSDB; ADK46098.
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
XX preventing and treating pathological conditions resulting from bacterial
XX infection, e.g. Streptococcus pneumoniae infection, and in drug
XX screening.
XX
XX Disclosure; SEQ ID NO 5274; 301pp; English.
XX
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
XX and polypeptides. The nucleic acids and proteins are useful for
XX diagnosing, preventing and treating pathological conditions resulting
XX from bacterial infection, such as S. pneumoniae infection. These may also
XX be used for drug screening procedures. The present sequence represents a
XX Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
XX data for this patent did not appear in the printed specification but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
XX
XX Query Match 99.6%; Score 845; DB 8; Length 2138;
XX Best Local Similarity 99.4%; Pred. No. 2.3e-73;
XX Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TTVKPILNKTGEVSELKPHRVTTVIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDG 60
XX DB 1941 TTVKPEILNKTGEVSELKPHRVTTVIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDG 2000
XX
XX QY 61 WEISGPEGKDGAGYVNLNLSKDTFIKVPFKIEEKKKEENKPTFDVSKKQNPQVNHSQLN 120
XX DB 2001 WEISGPEGKDGAGYVNLNLSKDTFIKVPFKIEEKKKEENKPTFDVSKKQNPQVNHSQLN 2060
XX
XX QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
XX DB 2061 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104
XX
XX RESULT 10
XX AAW55096
XX ID AAW55096 standard; protein; 117 AA.
XX
XX AC AAW55096;
XX
XX 02-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae SP0043 protein.
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
XX WPI; 1998-272224/24.
XX
XX N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX - or their epitope-containing fragments, useful in protective or
XX therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
XX
XX Query Match 72.5%; Score 615; DB 2; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 48 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLNLSKDTFIKVPFKIEEKKKEENKPTFDVSK 107
XX DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLNLSKDTFIKVPFKIEEKKKEENKPTFDVSK 60
XX
XX QY 108 KQNPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
XX DB 61 KQNPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 11
XX ABP54590
XX ID ABP54590 standard; protein; 117 AA.
XX
XX AC ABP54590;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
```

```

XX PD 23-MAY-2002.
XX PF 22-JAN-2001; 2001US-00765272.
XX XX 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
XX Query Match 72.5%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 48 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 107
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 60
QY 108 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 164
Db 61 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117
XX RESULT 12
XX ADC45149
XX ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX US6573082-B1.
XX PN 03-JUN-2003.
XX XX 28-MAR-2000; 2000US-00536784.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR

```

```

PR 30-OCT-1997; 97US-00961083.
XX XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX XX
XX PS Example 1; SEQ ID NO 68; 56pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
XX Query Match 72.5%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 48 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 107
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 60
QY 108 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 164
Db 61 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117
XX RESULT 13
XX AAG81779
XX ID AAG81779 standard; protein; 746 AA.
XX AC AAG81779;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX KW endocarditis.
XX OS Staphylococcus epidermidis.
XX XX WO200134809-A2.
XX PN 17-MAY-2001.
XX PD 09-NOV-2000; 2000WO-US030782.
XX PF 09-NOV-1999; 99US-0164258P.
XX PR (GLAX ) GLAXO GROUP LTD.
XX PA Kimmerly WJ;
XX PI WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.

```



PT New isolated polypeptide and encoding nucleic acid derived from  
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
 PT treating an S. epidermidis bacterial infection.

PS Claim 17; SEQ ID NO 5663; 741pp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for S. epidermidis infection; a recombinant or substantially  
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the Staphylococcus genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the Staphylococcus plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the Staphylococcus  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the Staphylococcus genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial  
 CC infection. This is the amino acid sequence of a S. epidermis protein of  
 CC the invention.

XX SQ Sequence 778 AA;

Query Match 14.0%; Score 119; DB 8; Length 778;  
 Best Local Similarity 27.0%; Pred. No. 0.019;  
 Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
 QY 10 KDTGEVSELKPHRVTVVTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW-- 61  
 Db 617 EDSVNAQSLKP-----ITIGNGKQIKQOQSVKSTKVLPHSKVLMVTDGELTMP-DMTGWTK 671  
 QY 62 -RISGFE-----GKQDAGYVIN--LSKDTPIKPVFKKIEBKKEENKPTPDVS-----K 107  
 Db 672 EDVLAPELTKIKVSTKNGFVTNQISKGQIILK-----NKKIEVLSLSARD 718  
 QY 108 KDNQVNHSQLNESHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNN 161  
 Db 719 TDDQEKTDDESDSNKSKDKADEHDSNTSSSTKN-----DKSNADSKNDSD 766

RESULT 16

ADZ72253  
 ID ADZ72253 standard; protein; 354 AA.

XX AC ADZ72253;

XX DT 14-JUL-2005 (first entry)

XX DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.

XX KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
 KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.

XX OS Plasmodium falciparum.

XX PN EP1526178-A1.

XX PD 27-APR-2005.

XX PF 24-OCT-2003; 2003EP-00292673.

XX

PR 24-OCT-2003; 2003EP-00292673.

XX PA (INSP ) INST PASTEUR.

XX PI Druilhe P;

XX WPI; 2005-323987/34.

DR N-PSDB; ADZ72252.

XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
 PT falciparum, which encode proteins useful for preparing vaccine  
 PT compositions against malaria.

XX Disclosure; SEQ ID NO 2; 137pp; English.

XX The present invention relates to the protection against malaria. More  
 CC particularly, the invention pertains to a family of MSP-3 (merozoite  
 CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
 CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
 CC falciparum, highly conserved in P. falciparum strains, simultaneously  
 CC expressed in P. falciparum at the erythrocytic stages and encoding  
 CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
 CC their N-terminal extremity and which are located at the merozoite  
 CC surface. The characterization of this gene family enables the definition  
 CC of immunogenic and vaccine compositions against P. falciparum. The  
 CC present sequence is the P. falciparum MSP-3-1 protein.

XX SQ Sequence 354 AA;

Query Match 13.1%; Score 111.5; DB 9; Length 354;  
 Best Local Similarity 22.5%; Pred. No. 0.035;  
 Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

QY 1 TTVKSEFILN-KDTGEVSELKPHRVTVVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFD 59  
 Db 154 TKTYEAKAKNAVEKAKNAQYQKAVLKAKKES-----YDYIL----- 194  
 QY 60 GWEISGP--EGKQDAG-----YVINLSKDTPIKPVFKKIEBKKE-----EN 99  
 Db 195 GWEFGGVPPEHKKEENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEEBELEKKEE 254  
 QY 100 KPTFDVSKKNQPNVNSQLNESHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTT 159  
 Db 255 ETESEISDEBEKEEKEEENDKCKEKEQENNNNDQKQMEAE-----QNLISKQNN 309  
 QY 160 NN 161  
 Db 310 NN 311

RESULT 17

ABU42797  
 ID ABU42797 standard; protein; 775 AA.

XX AC ABU42797;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #28324.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus epidermidis.

XX PN WO20027183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.







Query Match 13.0%; Score 110.5; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.1;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 4 KEFILNKOTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53  
Db 422 EBAVSEKNAHETVE---HEETVQSNEPKADNGVSNQNNELNENEFV-----ESE 472  
QY 54 K-----GYQFD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIE 92  
Db 473 KSEHEARSKAKESASYDYILGWFGGVPKHEENMLSHLYVSSKDKENISKENDVDLD 532  
QY 93 EKKEEENKPTFDVSKKDNQPNVHSQLN-----ESHKEDLQREHSQKSDS 139  
Db 533 E-KEEAETEEELKEEKEETSEISEDEEEEEKEEENEKKEQKEQSNENNDQ 591  
QY 140 TKDVTATVLDKNNISSKSTNN 161  
Db 592 KKDMA-----QNLISKQNNN 608

RESULT 21  
ADO19010  
ID ADO19010 standard; protein; 651 AA.  
XX  
AC ADO19010;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE P. falciparum GLURP-MSP3 fusion protein.  
XX  
KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
KW antimalarial; immunostimulant.  
XX  
OS Plasmodium falciparum.  
OS Synthetic.  
XX  
PN WO2004043488-A1.  
XX  
PD 27-MAY-2004.  
XX  
PF 06-NOV-2003; 2003WO-DK0000759.  
XX  
PR 12-NOV-2002; 2002DK-00001741.  
PR 11-SEP-2003; 2003DK-00001307.  
XX  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Theisen M, Jepsen S;  
XX  
DR WPI; 2004-411650/38.  
DR N-PSDB; ADO19011.  
XX  
PT New antigen based vaccine comprising a fusion protein derived from  
PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
PT preventing malaria.  
XX  
PS Claim 5; SEQ ID NO 1; 52pp; English.  
XX  
CC The present invention relates to a fusion protein comprising Plasmodium  
CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
CC useful as an antigen based vaccine against malaria. Also disclosed is the  
CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
CC polynucleotide sequence is also useful in preparing a vaccine. The  
CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 fusion protein.  
XX  
SQ Sequence 651 AA;  
Query Match 13.0%; Score 110.5; DB 8; Length 651;

Best Local Similarity 22.8%; Pred. No. 0.1;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 4 KEFILNKOTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53  
Db 422 EBAVSEKNAHETVE---HEETVQSNEPKADNGVSNQNNELNENEFV-----ESE 472  
QY 54 K-----GYQFD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIE 92  
Db 473 KSEHEARSKAKESASYDYILGWFGGVPKHEENMLSHLYVSSKDKENISKENDVDLD 532  
QY 93 EKKEEENKPTFDVSKKDNQPNVHSQLN-----ESHKEDLQREHSQKSDS 139  
Db 533 E-KEEAETEEELKEEKEETSEISEDEEEEEKEEENEKKEQKEQSNENNDQ 591  
QY 140 TKDVTATVLDKNNISSKSTNN 161  
Db 592 KKDMA-----QNLISKQNNN 608

RESULT 22  
ADZ79639  
ID ADZ79639 standard; protein; 188 AA.  
XX  
AC ADZ79639;  
XX  
DT 14-JUL-2005 (first entry)  
XX  
DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.  
XX  
KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
KW immunotherapy; malaria; antimalarial; vaccine.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO2005040206-A1.  
XX  
PD 06-MAY-2005.  
XX  
PF 22-OCT-2004; 2004WO-EP012910.  
XX  
PR 24-OCT-2003; 2003US-00691672.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Druilhe P;  
XX  
DR WPI; 2005-355821/36.  
XX  
CC Chimeric molecule useful for preparing vaccine composition against  
CC malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
CC protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
CC immunized with molecule.  
XX  
PS Disclosure; SEQ ID NO 7; 79pp; English.  
XX  
CC The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID NO:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID NO:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP

CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
CC MSP3f fragment. Note: The present sequence given as SEQ ID NO:7 in the  
CC Sequence Listing is not mentioned elsewhere in the specification.  
XX  
SQ Sequence 188 AA;

Query Match 12.7%; Score 108; DB 9; Length 188;  
Best Local Similarity 23.2%; Pred. NO. 0.032;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 27 IQNGKEMSTIVSEEDFILPVYKGELEKGYQFGWEISGF--EGKKDAG-----VYINLS 79  
DB 15 VLKAKEASS-----YDYL-----GWEPGGVPEHKKENMLSHLYVSKD 55  
QY 80 KDTFKVPFKIEEKKE-----ENKPTFVSKKKDPQVNHVSQINSHRKE 126  
DB 56 KENISKENDVDLDEKEEAETEEBELEKNEETSEISEDEBEBEKEEKEENDKKK 115  
QY 127 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 161  
DB 116 EQEKEQSNENNDQKKMEA-----QNLISKQNNN 145

RESULT 23  
ADT56185  
ID ADT56185 standard; protein; 470 AA.

XX AC ADT56185;  
XX  
XX DT 13-JAN-2005 (first entry)  
XX DE Plant polypeptide, SEQ ID 6262.  
XX KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
XX KW disease resistance; galactomannan production; plant growth regulator;  
XX KW heat tolerance; herbicide tolerance; lignin production;  
XX KW extreme osmotic condition tolerance; pathogens resistance;  
XX KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX OS Viridiplantae.  
XX  
XX PN US2004216190-A1.  
XX PD 28-OCT-2004.  
XX PP 18-DEC-2003; 2003US-00739930.  
XX PR 28-APR-2003; 2003US-00424599.  
XX PR 28-APR-2003; 2003US-00425115.  
XX PA (KOVA/) KOVALIC D K.  
XX  
XX PI Kovalic DK;  
XX  
XX DR WPI; 2004-757369/74.  
XX  
XX PT New recombinant DNA constructs useful in the field of biochemistry and  
XX PT genetics, and in particular for producing transgenic plants with improved  
XX PT biological characteristics.  
XX  
XX PS Claim 2; SEQ ID NO 6262; 14pp; English.  
XX  
XX CC The invention relates a recombinant DNA construct comprising a  
XX CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
XX CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
XX CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
XX CC Arabidopsis, wheat and rape but the specification does not indicate which  
XX CC sequences is derived from which organism. Also included is a method of  
XX CC producing a plant having an improved property, comprising transforming a  
XX CC plant with a recombinant DNA construct comprising a promoter region  
XX CC functional in a plant cell operably joined to a polynucleotide encoding a  
XX CC polypeptide associated with the property, and growing the transformed

CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
XX SQ Sequence 470 AA;

Query Match 12.7%; Score 107.5; DB 8; Length 470;  
Best Local Similarity 20.1%; Pred. NO. 0.13;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 19 KPHEVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFGWEISGF-----GKK 70  
DB 82 RENRVTDTVQNNNGESK-----YVODLARRIRYDE-EATGSQAQRIDHPNOK 129  
QY 71 DAGVYINLSKDTFKVPFKIEEKKEENKPTFVSKKKDN----- 111  
DB 130 NVGITERAFENSPIEETSHRVDDNKRINNQNFTAAKSENNAVSRVSGADHKRAEVMGK 189  
QY 112 PQVNHVSQIN-----SHRKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 164  
DB 190 PMENRDQVRQTESAKSHRKENVTSEKPRDQEGVKTEAKDKDRNKKKEKTESINK 248

RESULT 24  
AAG47777  
ID AAG47777 standard; protein; 484 AA.

XX AC AAG47777;  
XX  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PP 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145152P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 16-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149436P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.







	XX PF	21-MAR-2002; 2002WO-US009107.	
	XX PR	21-MAR-2001; 2001US-00815242.	
	PR PR	06-SRP-2001; 2001US-00948993.	
	PR PR	25-OCT-2001; 2001US-0342923P.	
	PR DR	08-FEB-2002; 2002US-000728S1.	
	PR XX	06-MAR-2002; 2002US-0362699P.	
	PA PA	(ELIT-) ELITRA PHARM INC.	
	PI PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;	
	PI PL	Mall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
	XX DR	WI; 2003-029926/02.	
	DR DR	N-PSDB; ACA2827A.	
	XX PT	New antitense nucleic acids, useful for identifying proteins or screening	
	PT PT	for homologous nucleic acids required for cellular proliferation to	
	PT PT	isolate candidate molecules for rational drug discovery programs.	
	XX PS	Claim 25; SEQ ID NO 52328; 1766pp; English.	
	XX CC	The invention relates to an isolated nucleic acid comprising any one of	
	CC CC	the 6213 antisense sequences given in the specification where expression	
	CC CC	of the nucleic acid inhibits proliferation of a cell. Also included are:	
	CC CC	(1) a vector comprising a promoter operably linked to the nucleic acid	
	CC CC	encoding a polypeptide whose expression is inhibited by the antisense	
	CC CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated	
	CC CC	polypeptide or its fragment whose expression is inhibited by the	
	CC CC	antisense nucleic acid; (4) an antibody capable of specifically binding	
	CC CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
	CC CC	proliferation or the activity of a gene in an operon required for	
	CC CC	proliferation; (7) identifying a compound that influences the activity of	
	CC CC	the gene product or that has an activity against a biological pathway	
	CC CC	required for proliferation, or that inhibits cellular proliferation; (8)	
	CC CC	identifying a gene required for cellular proliferation or the biological	
	CC CC	pathway in which a proliferation-required gene or its gene product lies	
	CC CC	or a gene on which the test compound that inhibits proliferation of an	
	CC CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a	
	CC CC	compound's activity; (11) a culture comprising strains in which the gene	
	CC CC	product is overexpressed or underexpressed; (12) determining the extent	
	CC CC	to which each of the strains is present in a culture or collection of	
	CC CC	strains; or (13) identifying the target of a compound that inhibits the	
	CC CC	proliferation of an organism. The antisense nucleic acids are useful for	
	CC CC	identifying proteins or screening for homologous nucleic acids required	
	CC CC	for cellular proliferation to isolate candidate molecules for rational	
	CC CC	drug discovery programs, or for screening homologous nucleic acids	
	CC CC	required for proliferation in cells other than S. aureus, S. typhimurium,	
	CC CC	K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of	
	CC CC	the target prokaryotic essential genes. Note: The sequence data for this	
	CC CC	patent did not form part of the printed specification, but was obtained	
	CC CC	in electronic format directly from WIPO at	
	CC CC	ftp.wipo.int/pub/published_pct_sequences	
	XX SQ	Sequence 903 AA;	
		Query Match            12.1%; Score 103; DB 6; Length 903;	
		Best Local Similarity     21.9%; Pred. No. 0.88;	
		Matches      42; Conservative    37; Mismatches    59; Indels    54; Gaps        8	
	QY	7 ILNKDTGEVSELKPHRVTVTTIQNGEMSSSTIVSEDFILPVYKGELKEGYQPDGWEISGP 66	
	Dd	:   :::  : :    ::  :: :   ::     :	
	Db	502 VLNRNTQLNVEKS KOLPKINVELKEERKNQIIRKH-----ELEKG----- 542	
	QY	67 EGKDAGGVYNLSKDTFIKPVKCIIEKKBEENKFTFDS-----KKDONPNVN----- 115	
	Dd	::  : :  : :  : :  : :  : :  : :  : :  : :	
	Db	543 KGQPDPKVNIELNKEK-VKQVFNEISESKNEKSQQVVNRNTQLNVEKS KOLPKINVELKE 601	
	QY	116 -----HSQLNESHRKE-----DLQR-----EHSQS-KSDSTKDVATVLDKNN 152	
	Dd	:: :: :  : :  : :  : :  : :  : :  : :  : :	
	Db	602 EKSQTSEKHNLREKRGRKIPKANVELLVEKS QIIKEHNLEKESKTQTPKWVNLNK 661	
	Dd	:	

```

Qy      153 ISSKSTTNPNK 164
       :| | |
Db      662 EKAKHFNESIK 673

RESULT 29
ABB61977
XX AC ABB61977 standard; protein; 564 AA.
XX AC ABB61977;
XX DT 26-MAR-2002 (first entry)
XX XX Drosophila melanogaster polypeptide SEQ ID NO 12723.
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL06080.
XX FT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX PS Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLU6176-ABLU3051), expressed DNA
CC sequences (ABLU01840-ABLU6175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 564 AA;

Query Match          12.0%; Score 101.5; DB 4; Length 564;
Best Local Similarity 24.5%; Pred.No. 0.65;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

Qy      41 EDFILPVYKGELBKGVPDGV-----RISGFEGKTDAGVI-----NLSKDTFIK 85
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      78 EDLDTPLBSRFPSK--VFDCWVDHRDEHDGHDVQPSGEALDDHDDHDDHDEDEE 135

Qy      86 PVFKCLEEKKEENKPT-----FDYSKKDNQPVNHSQLNESHRXEDLQREHSQSDDS 139
       ||:::||:|||||:| | | | | | | | | | | | | | | | | | | | | |
Db      136 PLTEEEELLEEBEPTEDESPAADBEYEDEDENNA--GENITADAEERSEBDND 193

Qy      140 TKDVTATVLDKNNISSKST 158
       |||||
Db      194 EGTVEATVEATTTEAT 212
       |||||

RESULT 30

```



PT cancer or hepatitis.

PS Claim 1; Page 207-211; 238pp; English.

XX

XX The present invention describes 23 human molecules for disease detection

CC and treatment (MDDT-1 to 23) (see ABP55392 to ABP55414). The human MDDT-1

CC to 23 proteins (I) are encoded by the sequences given in ABQ3859 to

CC ABQ3881. (I) can have various activities depending on the cells and

CC tissues in which they are expressed. These activities include: anti-HIV;

CC anti-allergic; anti-inflammatory; antianaemic; antiparkinsonian; nootropic;

CC anticonvulsant; antiinfertility; antiarteriosclerotic; antiasthmatic;

CC immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;

CC antiidiabetic; nephrotropic; antigitout; thyromimetic; neuroprotective;

CC osteopathic; antiarthritic; antiparasitic; haemostatic; antibacterial;

CC uropathic; protozoacide; and fungicide. (I) and the polynucleotides

CC encoding them can be used in gene therapy. (I), polynucleotides, agonists

CC and antagonists from the present invention can be used for diagnosing, or

CC treating or preventing disorders associated with aberrant expression of

CC MDDT, particularly cell proliferative (e.g. cancer), developmental

CC disorders, neurological disorders, reproductive disorders, or

CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,

CC parasitic, protozoal or helminthic infections. They are also useful in

CC the assessment of the effects of exogenous compounds on the expression of

CC nucleic acid and amino acid sequences of proteins associated with MDDT

XX

SQ Sequence 1384 AA;

Query Match 11.7%; Score 99.5; DB 6; Length 1384;

Best Local Similarity 21.7%; Pred. No. 3.5;

Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

QY 4 KEFILNKDTGVESELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQF----- 58

DB 1013 KEFILNQEQEISQLK-KEIERTQQRKEMESVMKEQEQYIATQYKKAIDLGQELRLTR 1071

QY 59 -----DGWEISGFEGKKDA-----GVVINLSK----- 80

DB 1072 QVQNSHTLAEARHQVQAQREIERLSSELDMMKQLSKEDAGHNLAEELGASKVREAH 1131

QY 81 -----DTPIKPVFKLEKKE-----EENKPTFDVSKKKNQPNVHNSQLNESHKRE 126

DB 1132 LEARMQAEIKKLAEVESLKEAYHMEMISHQENHAKWKIS--ADSQKSSVQQLNQLERKA 1189

QY 127 DLQREHHSQKSDSTKDVTTATVLDKNNI 153

DB 1190 KLELEE---AQDTVSNLHQVQDRNEV 1213

RESULT 32

ABP55393

ID ABP55393 standard; protein; 1404 AA.

XX

AC ABP55393;

XX

DT 04-FEB-2003 (first entry)

XX

DE Human MDDT-2 protein SEQ ID NO:2.

XX

XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;

KW anti-allergic; anti-inflammatory; antianaemic; antiparkinsonian; nootropic;

KW anticonvulsant; antiinfertility; antiarteriosclerotic; antiasthmatic;

KW immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;

KW antiidiabetic; nephrotropic; antigitout; thyromimetic; neuroprotective;

KW osteopathic; antiarthritic; antiparasitic; haemostatic; antibacterial;

KW uropathic; ophthalmological; antiparasitic; haemostatic; antibacterial;

KW virucide; protozoacide; fungicide; gene therapy; cell proliferative;

KW cancer; developmental disorder; neurological disorder; infection;

KW reproductive disorder; autoimmune disorder; inflammatory disorder.

XX

OS Homo sapiens.

XX

PN WO200278420-A2.

XX 10-OCT-2002.

XX

PF 29-MAR-2002; 2002WO-US009809.

XX

XX 30-MAR-2001; 2001US-0280387P.

PR 05-APR-2001; 2001US-0282335P.

PR 13-APR-2001; 2001US-0283663P.

PR 18-APR-2001; 2001US-0285484P.

PR 18-JAN-2002; 2002US-0350702P.

PR 25-JAN-2002; 2002US-0351749P.

XX (INCY-) INCYTE GENOMICS INC.

PA

XX Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;

PI Ramkumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY, Warren BA;

PI Nguyen DB, Thanagavelu K, Yao MG, Elliott VS, Baughn MR, Emerling BM;

PI Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;

XX

DR WPI; 2003-058385/05.

DR N-PSDB; ABQ3860.

XX

XX New human molecules for disease detection and treatment, useful for

PT diagnosing, treating or preventing autoimmune or inflammatory disorders

PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,

PT cancer or hepatitis.

XX

PS Claim 1; Page 172-175; 238pp; English.

XX

CC The present invention describes 23 human molecules for disease detection

CC and treatment (MDDT-1 to 23) (see ABP55392 to ABP55414). The human MDDT-1

CC to 23 proteins (I) are encoded by the sequences given in ABQ3859 to

CC ABQ3881. (I) can have various activities depending on the cells and

CC tissues in which they are expressed. These activities include: anti-HIV;

CC anti-allergic; anti-inflammatory; antianaemic; antiparkinsonian; nootropic;

CC anticonvulsant; antiinfertility; antiarteriosclerotic; antiasthmatic;

CC immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;

CC antiidiabetic; nephrotropic; antigitout; thyromimetic; neuroprotective;

CC osteopathic; antiarthritic; antiparasitic; haemostatic; antibacterial;

CC uropathic; ophthalmological; antiparasitic; haemostatic; antibacterial;

CC virucide; protozoacide; and fungicide. (I) and the polynucleotides

CC encoding them can be used in gene therapy. (I), polynucleotides, agonists

CC and antagonists from the present invention can be used for diagnosing,

CC treating or preventing disorders associated with aberrant expression of

CC MDDT, particularly cell proliferative (e.g. cancer), developmental

CC disorders, neurological disorders, reproductive disorders, or

CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,

CC parasitic, protozoal or helminthic infections. They are also useful in

CC the assessment of the effects of exogenous compounds on the expression of

CC nucleic acid and amino acid sequences of proteins associated with MDDT

XX

SQ Sequence 1404 AA;

Query Match 11.7%; Score 99.5; DB 6; Length 1404;

Best Local Similarity 21.7%; Pred. No. 3.6;

Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

QY 4 KEFILNKDTGVESELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQF----- 58

DB 1033 KEFILNQEQEISQLK-KEIERTQQRKEMESVMKEQEQYIATQYKKAIDLGQELRLTR 1091

QY 59 -----DGWEISGFEGKKDA-----GVVINLSK----- 80

DB 1092 QVQNSHTLAEARHQVQAQREIERLSSELDMMKQLSKEDAGHNLAEELGASKVREAH 1151

QY 81 -----DTPIKPVFKLEKKE-----EENKPTFDVSKKKNQPNVHNSQLNESHKRE 126

DB 1152 LEARMQAEIKKLAEVESLKEAYHMEMISHQENHAKWKIS--ADSQKSSVQQLNQLERKA 1209

QY 127 DLQREHHSQKSDSTKDVTTATVLDKNNI 153

DB 1210 KLELEE---AQDTVSNLHQVQDRNEV 1233





CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1184 AA;

Query Match 11.5%; Score 97.5; DB 6; Length 1184;  
Best Local Similarity 26.3%; Pred. No. 4.5;  
Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
QY 14 EVSELKPHRVTTIQ-NGKEMSTIV--SEDP--ILPVYKGLSKGVQPDGWEISGREG 68  
Db 240 ELSEVNEHRKVIKELNEKEQKNVVEKKQEDINKEVEVLQDVIEKSVDIYN-SIKGVIS 298  
QY 69 KKDAGVIVNLSKDTF-----IKPVKKIEEKK-----EENKPTFDVSKKD 110  
Db 299 KES--QINLIERIRNFTNEISRKNLEIKDKEKLNENKQYKLESNK--LSGSEELS 354  
QY 111 NPQVNHSQLNESHKEDLQREBHSQKSDSTKDVATVLD-KNNISKSTTNNPNK 164  
Db 355 TLQENIKVLEGSKDQKQIKGLSINNEIELLKESIIDLNNKQDFSNKLSLTNANK 409

RESULT 37  
ABR64281  
ID ABR64281 standard; protein; 2468 AA.  
XX  
AC ABR64281;  
XX  
DT 15-OCT-2003 (first entry)  
XX  
DE Angiogenesis protein BNO382.  
XX  
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic; angiogenesis;  
KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
KW diabetic retinopathy; cardiovascular disease; atherosclerosis;  
KW ischemic limb disease; coronary artery disease.  
XX  
OS Homo sapiens.  
XX  
PN WO2003027285-A1.  
XX  
PD 03-APR-2003.  
XX  
PF 19-SEP-2002; 2002WO-AU001282.  
XX  
PR 27-SEP-2001; 2001AU-00007973.  
PR 27-SEP-2001; 2001AU-00007974.  
PR 11-OCT-2001; 2001AU-00008210.  
PR 29-OCT-2001; 2001AU-00008532.  
PR 13-NOV-2001; 2001AU-00008838.  
PR 28-AUG-2002; 2002AU-00951032.  
XX  
PA (BION-) BIONOMICS LTD.  
XX  
PI Gamble JR, Hahn CN, Vadas MA;  
XX  
XX WPI; 2003-354655/33.  
DR N-PSDB; ACF34559.  
XX  
XX New angiogenic genes and polypeptides, useful for diagnosing,  
PT prognosticating or treating an angiogenesis-related disorder, e.g.  
PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or  
XX cardiovascular diseases.  
XX  
PS Claim 15; SEQ ID NO 216; 90pp; English.  
XX

CC The invention relates to the isolation of novel genes (ACF34446-ACF34559)  
CC encoding proteins (ABR64180-ABR64281) involved in the process of  
CC angiogenesis. The nucleic acid molecules are useful in identifying and/or  
CC obtaining full-length human genes involved in an angiogenic process. The  
CC nucleic acid molecule, polypeptides or complexes encoded, cells or  
CC genetically modified non-human animals derived from these are useful for  
CC the screening of candidate pharmaceutical compounds used in treating  
CC angiogenesis-related disorders. They are also useful for diagnosing,  
CC prognosticating or treating an angiogenesis-related disorder, which  
CC involves uncontrolled or enhanced angiogenesis or is a disorder in which  
CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,  
CC diabetic retinopathy, psoriasis or cardiovascular diseases such as  
CC atherosclerosis), or involves inappropriately arrested or decreased  
CC angiogenesis or is a disorder in which an expanding vasculature is of  
CC benefit (e.g. ischemic limb disease or coronary artery disease). The  
CC modulator of expression or activity of the polypeptide encoded by the  
CC nucleic acid sequence is useful for manufacturing a medicament for the  
CC treatment of an angiogenesis-related disorder. This sequence corresponds  
CC to one of the novel angiogenic protein  
XX  
SQ Sequence 2468 AA;

Query Match 11.5%; Score 97.5; DB 6; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;  
QY 4 KEFILNKDTGVSSELKPHRVTTIQKEMSTIVSEDFILPVYKGLSKGVQPDGWEI 63  
Db 584 EKVMVKKDKPVKTKTSPVTEKVPSEEPS-----PV-KAEVA-----EK 623  
QY 64 SGFEGKKDAGVIVNLSKDTFIKPVFKIEEKKEBENKPTPDVSKKONPQVNHSQLNESH 123  
Db 624 QATDVKPKAAKEKTVKKTVP-----EDKKEEKKPKKEVAKEDKTPi---KKEKP 675  
QY 124 RKEDLQRE-----BHSQKSDSTKDV 143  
Db 676 KKEEVKKEVKKEIKKEEKKPKKEV 700  
RESULT 38  
ADE62723  
ID ADE62723 standard; protein; 2468 AA.  
XX  
AC ADE62723;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein NP\_005900, SEQ ID NO 8656.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO) GEN HOSPITAL CORP.  
XX (FARB) BAYER AG.  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
DR GENBANK; NP\_005900.  
XX









GenCore version 5.1.17  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 75.3154 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 BDFILPVYKGLKGYQFDG.....ATVLDKNNISSKSTTNNPK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	651	100.0	2 Q9AHT5_STRPN	Q9Aht5 streptococc
2	651	100.0	2 Q97RY6_STRPN	Q97ry6 streptococc
3	648	99.5	2 Q9S4M8_STRPN	Q9S4m8 streptococc
4	648	99.5	2 Q8DQ7_STRPN	Q8dqp7 streptococc
5	125.5	19.3	2 Q4XUI6_PLACH	Q4xui6 plasmodium
6	112	17.2	2 Q95P15_PLAFA	Q95p15 plasmodium
7	110	16.9	2 Q9U0G0_PLARE	Q9u0g0 plasmodium
8	110	16.9	2 Q25705_PLAFA	Q25705 plasmodium
9	109	16.7	2 Q8I436_PLAF7	Q8i436 plasmodium
10	108	16.6	2 Q5OVJ0_ENTHI	Q5ovj0 entamoeba h
11	106.5	16.4	2 Q77355_PLAF7	Q77355 plasmodium
12	106	16.3	2 Q25995_PLAFA	Q25995 plasmodium
13	106	16.3	2 Q8IJ55_PLAF7	Q8ij55 plasmodium
14	106	16.3	2 Q6BRW2_DEBHA	Q6brw2 debaromyce
15	103	15.8	2 Q9U6C4_PLAFA	Q9u6c4 plasmodium
16	102.5	15.7	2 Q26019_PLAFA	Q26019 plasmodium
17	102	15.7	2 Q25706_PLAFA	Q25706 plasmodium
18	102	15.7	2 Q4WTJ2_DICDI	Q4wtj2 dictyosteli
19	101.5	15.6	2 Q9V7J0_DROME	Q9v7j0 drosophila
20	101.5	15.6	2 Q9V7I9_DROME	Q9v7i9 drosophila
21	101.5	15.6	2 Q9GQ82_DROME	Q9gq82 drosophila
22	101.5	15.6	2 Q6HNR0_BACHK	Q6hnr0 bacillus th
23	101	15.5	2 Q50LX8_ENTHI	Q50lx8 entamoeba h
24	101	15.5	2 Q4TMO4_PLARE	Q4tmo4 plasmodium
25	100	15.4	2 Q4Y2I3_PLACH	Q4y2i3 plasmodium
26	100	15.4	2 Q8IJZ4_PLAF7	Q8ijz4 plasmodium
27	99.5	15.3	2 Q90784_CHICK	Q90784 gallus gall
28	99	15.2	2 P91488_CAENOR	P91488 caenorhabdi
29	99	15.2	2 Q9NFV9_PLAFA	Q9nfv9 plasmodium
30	99	15.2	2 Q54K26_DICDI	Q54k26 dictyosteli
31	99	15.2	2 Q8I3A0_PLAF7	Q8i3a0 plasmodium

32	98.5	15.1	393	2 Q7RKU2_PLAYO	Q7rku2 plasmodium
33	98.5	15.1	674	2 Q7RLI7_PLAYO	Q7rle7 plasmodium
34	98.5	15.1	827	2 Q55G46_DICDI	Q55g46 dictyosteli
35	98	15.1	951	2 Q96229_PLAF7	Q96229 plasmodium
36	97.5	15.0	556	2 Q9S893_DROME	Q9s893 drosophila
37	97.5	15.0	1345	1 YH00_YEAST	P38800 saccharomyc
38	97	14.9	895	2 Q4Z7T5_PLABE	Q4z7t5 plasmodium
39	97	14.9	1028	2 Q4Z4Q1_PLABE	Q4z4q1 plasmodium
40	96.5	14.8	905	2 Q6BXE1_DEBHA	Q6bxel debaromyce
41	96	14.7	736	2 Q4YVY2_PLABE	Q4yvy2 plasmodium
42	96	14.7	899	2 Q86J06_DICDI	Q86jq6 dictyosteli
43	96	14.7	1042	2 Q7REQ7_PLAYO	Q7req7 plasmodium
44	96	14.7	1120	2 Q557C8_DICDI	Q557c8 dictyosteli
45	95.5	14.7	500	2 Q6BGL7_PARTE	Q6bgl7 paramecium

## ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S;  
RA "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598 (2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08\_064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR018999; Gram\_pos\_anchor.  
DR InterPro; IPR00137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPTXG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match
 100.0%; Score 651; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9.9e-41;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 60
D 1962 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 201
D 1962 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 201

QY 61 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 120
D 2022 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2081
D 2022 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2081

QY 121 NPNK 124
D 2082 NPNK 2085

RESULT 2
Q37RY6 STRPN
ID Q37RY6 STRPN PRELIMINARY; PRT; 2140 AA.
AC Q37RY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocName=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin N., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.H., Heidelberg J.P., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollinghead S.K., Frazer C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF06280; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA4AD8E2938B334 CRC64;

Query Match
 100.0%; Score 651; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1e-40;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 60
D 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 2042
D 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVINLSKDTFKPVPFKIEEKEEENK 2042

QY 61 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 120
D 2043 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2102
D 2043 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2102

QY 121 NPNK 124
D 2103 NPNK 2106

RESULT 3
Q9S4M8 STRPN
ID Q9S4M8 STRPN PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor PrtA.
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3.B;
RX MEDLINE=21585565; PubMed=11728722;
RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
RA Zysk G.;
RT "The cell wall-associated serine proteinase PrtA: a highly conserved virulence factor of Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 205:99-104(2001).
DR EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```

DR PROSITE, PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Cell wall; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 2144 cell wall-associated serine proteinase
FT PT
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match
Best Local Similarity 99.5%; Score 648; DB 2; Length 2144;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVNIISKDTFIKVPFKIEEKKKEENK 60
Db 1987 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVNIISKDTFIKVPFKIEEKKKEENK 2046

QY 61 PTFDVSKKKDPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN 120
Db 2047 PTFDVSKKKDPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN 2106

QY 121 NPNK 124
Db 2107 NPNK 2110

RESULT 4
Q8DQ7 STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQ7
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
DN Name=prtA; OrderedLocustNames=pr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA DeHoff B.S., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Hoskins J., Strem S.J., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AAK93365.1; -; Genomic_DNA.
DR FIR; A97942; A97942.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.

```

```

DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match
Best Local Similarity 99.5%; Score 648; DB 2; Length 2144;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVNIISKDTFIKVPFKIEEKKKEENK 60
Db 1987 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVNIISKDTFIKVPFKIEEKKKEENK 2046

QY 61 PTFDVSKKKDPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN 120
Db 2047 PTFDVSKKKDPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN 2106

QY 121 NPNK 124
Db 2107 NPNK 2110

RESULT 5
Q4XUI6 PLACH PRELIMINARY; PRT; 300 AA.
AC Q4XUI6;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
DN ORFNames=PC000286.03.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAATJ01003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 4.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 300 AA; 34469 MW; 8EDE4E512AFB1945 CRC64;

Query Match
Best Local Similarity 19.3%; Score 125.5; DB 2; Length 300;
Matches 33; Conservative 26; Mismatches 31; Indels 25; Gaps 6;

QY 5 LPVYKGLKGYQFDGWEISGFEGKDGAGYVNIISKDTFIKVPFKIEEKKKEENK 56

```

Db 164 LAVYRGHVPAYKI-AWSI-----DNNYIVSCSDSTLKLWLNHLVPLLRKENA 215  
 QY 57 ---EENKTFDVKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVL 108  
 Db 216 OTKDSQK-----NEQKENPQ-NDQPNDEANSEKCKKQKKEKNDKTKNKIKTL 264

## RESULT 6

Q95P15 PLAPA  
 ID Q95P15\_PLAPA PRELIMINARY; PRT; 361 AA.  
 AC Q95P15;  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Merozoite surface protein 3 (Fragment).  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN-PVO.  
 RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
 RA Hiseada H., Saul A., Reese J.J., Kennedy M.C., Long C.A., Miller L.H.,  
 RA Stowers A.W.;  
 RT "Merozoite surface protein 3 and protection against malaria in Aotus  
 RT nancymai monkeys";  
 RL J. Infect. Dis. 185:657-664 (2002).  
 DR EMBL; AY044180; AAK94780.1; -; Genomic\_DNA.  
 DR InterPro; IPR010784; Merozoite\_SPAM.  
 DR Pfam; PF07133; Merozoite\_SPAM; 1.  
 KW Merozoite.  
 FT CHAIN  
 FT NON\_TER <1 >361 merozoite surface protein 3.  
 FT NON\_TER 1  
 FT NON\_TER 361  
 FT SEQUENCE 361 AA; 41163 MW; 6127A3041587BA7A CRC64;

Query Match 17.2%; Score 112; DB 2; Length 361;  
 Best Local Similarity 23.9%; Pred. No. 1.5; Mismatches 48; Indels 42; Gaps 6;  
 Matches 37; Conservative 28;  
 QY 8 YKGELEKGYQ-----PD-----GWEISGF--EGKQDAG-----YVI 36  
 Db 165 YAGKVKDYERAKNAYQKANQAVLKAKESYDYILGWFGGVPEHKKEENMLSHLYS 224  
 QY 37 NLSKDTFTKPVFKKLEKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQK 96  
 Db 225 SKOKENISKENDVDLDE-KKEEABETEELKEEKNKEETSEISEDESEEEKEEKEE 283  
 QY 97 SDSTKDVATVLQKN-----NISKSTTN 121  
 Db 284 NDKKKEQEQSNENNDQKQMEAQNLISKQNNN 318

## Query Match 17.2%; Score 112; DB 2; Length 361;

Best Local Similarity 23.9%; Pred. No. 1.5; Mismatches 48; Indels 42; Gaps 6;  
 Matches 37; Conservative 28;  
 QY 8 YKGELEKGYQ-----PD-----GWEISGF--EGKQDAG-----YVI 36  
 Db 165 YAGKVKDYERAKNAYQKANQAVLKAKESYDYILGWFGGVPEHKKEENMLSHLYS 224  
 QY 37 NLSKDTFTKPVFKKLEKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQK 96  
 Db 225 SKOKENISKENDVDLDE-KKEEABETEELKEEKNKEETSEISEDESEEEKEEKEE 283  
 QY 97 SDSTKDVATVLQKN-----NISKSTTN 121  
 Db 284 NDKKKEQEQSNENNDQKQMEAQNLISKQNNN 318

## RESULT 7

Q9U0G0 PLARE  
 ID Q9U0G0\_PLARE PRELIMINARY; PRT; 346 AA.  
 AC Q9U0G0;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Merozoite surface protein 3 (Fragment).  
 GN Name=mep3;  
 OS Plasmodium reichenowi.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5854;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;  
 RA Okeno D.M.N., Thomas A.W., Conway D.J.;  
 RT "Allelic lineages of the merozoite surface protein 3 gene in  
 RT Plasmodium reichenowi and Plasmodium falciparum";  
 RL Mol. Biochem. Parasitol. 109:185-188 (2000).

DR EMBL; AJ252286; CAB65754.1; -; Genomic\_DNA.  
 DR InterPro; IPR010784; Merozoite\_SPAM.  
 DR Pfam; PF07133; Merozoite\_SPAM; 1.  
 KW Merozoite.  
 FT NON\_TER 1  
 FT NON\_TER 346  
 FT NON\_TER 346  
 FT SEQUENCE 346 AA; 39127 MW; AB04B96BDFAFAP010 CRC64;

## Query Match 16.9%; Score 110; DB 2; Length 346;

Best Local Similarity 26.2%; Pred. No. 2; Mismatches 54; Indels 16; Gaps 5;  
 Matches 34; Conservative 26;  
 QY 7 VYKGELEKGYQFO-GWEISGF--EGKQDAG-----YVINLSKDTFTKPVFKKIEKEE 58  
 Db 184 VLKAKESYNYILGWFGGVPEHKKEENMLSHLYSSKDKENISKENDVDLDE-KKEE 242  
 QY 59 NKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATV-----LDKN 111  
 Db 243 AEETGEQLEKEEKEETSEINEDEQEEBEEKEEENNDKKEQAKESQSNQDKEDMEAQ 302  
 QY 112 NISKSTTN 121  
 Db 303 NLISKQNNN 312

## RESULT 8

Q25705 PLAPA  
 ID Q25705\_PLAPA PRELIMINARY; PRT; 379 AA.  
 AC Q25705;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Polymorphic antigen.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
 RA McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
 RL Mol. Biochem. Parasitol. 90:21-31 (1997).  
 DR EMBL; U08851; AAC47831.1; -; Unassigned\_DNA.  
 DR InterPro; IPR010784; Merozoite\_SPAM.  
 DR Pfam; PF07133; Merozoite\_SPAM; 1.  
 DR SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

## Query Match 16.9%; Score 110; DB 2; Length 379;

Best Local Similarity 23.4%; Pred. No. 2.2; Mismatches 30; Indels 50; Gaps 7;  
 Matches 37; Conservative 30;  
 QY 8 YKGELEKGYQ-----PD-----GWEISGF--EGKQDAG-----YVI 36  
 Db 185 YAGKVKDYERAKNAYQKANQAVLKAKESYDYILGWFGGVPEHKKEENMLSHLYS 244  
 QY 37 NLSKDTFTKPVFKKLEKEENKPTFDVSKKQNPVNHSQLN-----ESH 83  
 Db 245 SKOKENISKENDVDLDE-KKEEABETEELKEEKNKEETSEISEDESEEEKEEKEE 303  
 QY 84 RKEDLQREHSQKSDSTKDVATVLDKNNISSTTN 121  
 Db 304 KKKQEQEQSNENNDQKQMEAQNLISKQNNN 336

## RESULT 9

Q8I436 PLAF7  
 ID Q8I436\_PLAF7 PRELIMINARY; PRT; 3008 AA.  
 AC Q8I436;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Hypothetical protein PFE0325w.

```

GN Name=PRE0325w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=3D7;
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3D7;
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrall B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929351; CAD51431.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBE15C599B4 CRC64;

Query Match 16.7%; Score 109; DB 2; Length 3008;
Best Local Similarity 32.4%; Pred. No. 24;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 34 YVINLSK----DTFKPVFKIEEKEEENKPTFDVSKKONPQVNHSHQ---LNESHRKE 86
Db 2310 YDIELSKLEKFGASIGPVFTD-EENKEENKN-EVNKEENKKEENKKEENKKE 2366

QY 87 DLQREH----SQSDSTKVTVATVLDKNNISK-----STNNPNK 124
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 10
Q50VJ0 ENTHI PRELIMINARY; PRT; 384 AA.
AC Q50VJ0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE HMG box protein.
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Lofius B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hart R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N.G., Gluchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,

RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrall B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF801000585; EAL45607.1; -; Genomic DNA.
SQ SEQUENCE 384 AA; 45464 MW; 207789F65D72B019 CRC64;

Query Match 16.6%; Score 108; DB 2; Length 384;
Best Local Similarity 32.0%; Pred. No. 3.1;
Matches 31; Conservative 18; Mismatches 42; Indels 6; Gaps 2;

QY 22 EISGFEGKDKAGYVINLSKDTFKIPVKIEEKEEENKPTFDVSKKONPQVNHSHQ 81
Db 268 EMKKNKGKESDKKEDTKDKKKVKKSEKDKIIEKDEKKH---EKKEEKEEKKPKPE 323

QY 82 SHRKEDLQREHSHQSOKSSTKD--VTATVLDKNNISK 116
Db 324 SEKESKKEKSHKSKEDKKDKDEKSKKVEDKSKKQK 360

RESULT 11
O77355 PLAF7 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein MAL3P4.20.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson S., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Horneby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., Mclean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream J.E.,
RA Rutter S., Sulston J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrall B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Lofius B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hart R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N.G., Gluchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,

```

```
DR InterPro; IPR002483; PMI.
DR Pfam; PF01480; PMI; 1.
DR SMART; SM00311; PMI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 16.4%; Score 106.5; DB 2; Length 600;
Best Local Similarity 29.4%; Pred. No. 6.6;
Matches 53; Conservative 21; Mismatches 45; Indels 61; Gaps 12;

QY 1 EDFILPVY-----KGELEKGYQFDGWEISGPEGKK-----DAGYVNLNLSKDTFKPVPFKKI 51
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 EDDILYECISQLKQKSKK---DGEEDKYLNAKKLKLNLGAFIGNKSDSIFIELLELL 120
QY 52 --EEKKEE-----ENKPTFDVSK-KDNQVNVHSQLNE-----SHRKE--- 86
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 INEEKKEEHIADTLNENK-TNDIKVKVKNENINENYVNEKNDISNKKDKEHVSHQNEHNI 179
QY 87 -----DLQREEH-----SQKSDSTK-----DVTATVLDKNNISSKSTTNNPNK 124
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 NNVNLKKEKEYTDIQDKKKHKSLSQKSDSYKKPFNKRKTSIER-SLSNKRVDKTNK 238

RESULT 12
Q25995 PLAF7
ID Q25995_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=9519874; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.P.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1; -; Genomic DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 16.3%; Score 106; DB 2; Length 354;
Best Local Similarity 23.5%; Pred. No. 4.1;
Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;

QY 7 VYKGELEKGYQFD-GWEISGF--EGKQDAG-----YVNLNLSKDTFKPVPFKKIEKKEE- 57
Db 181 VLKAKEASSDYILGWFGGVPFHKKEENMLSHLYVSSKDKENISKENDVDLDEKEEA 240
QY 58 -----ENKPTFDVSKKNQPNQVNHSQLNESHKEDLQREHKSQKSDSTKDVTA 105
Db 241 EETEEELKEKEEETEISEDEEEEEEEKEEENDKKKEQKEQSNENNDDQKQDMEA 300
QY 106 TVLDKNNISSKSTNN 121
Db 301 -----QNLISKQNNN 311

RESULT 14
Q6BRW2 DEBHA
ID Q6BRW2_DEBHA PRELIMINARY; PRT; 616 AA.
AC Q6BRW2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA4458|PF8464 Candida albicans IPF8464 unknown
DE function.
GN OrderedLocusNames=DEHA0014674g;
OS Debaryomyces hansenii (yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anhouart V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boiserie A., Boyer J.S., Cattolico L., Confanioli F., de Daruvar A.,
```





DR InterPro; IPR010784; Merozoite SPAM.  
DR Pfam; PF07133; Merozoite SPAM; 1.  
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54B1ED91A24 CRC64;

Query Match 15.7%; Score 102; DB 2; Length 379;  
Best Local Similarity 22.8%; Pred. No. 8.8;  
Matches 36; Conservative 30; Mismatches 42; Indels 50; Gaps 7;

QY 8 YKGELEKGYQ-----PD---GWEISGF--EGKKDAG-----YVI 36  
Db 185 YAEQVEKDYERAKNAVYQKANAQVILKAKBASSYDILYLGWFFGGVFEHKKENMLSHLYS 244

QY 37 NLSKQTFIKVPFKIEEKEENKPTFDVSKKDNQVNHSQLN-----ESH 83  
Db 245 SKDENISKENDVLDLDE-KEEAETETEELEKEEETETESISEDEEBEEREEKEEND 303

QY 84 RKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 121  
Db 304 KKKEQKEQSNENNQKKDMEA-----QNLISKNNQN 336

RESULT 18  
Q54MT2 DICDI PRELIMINARY; PRT; 540 AA.  
AC Q54MT2;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein.  
GN ORFNames=DDB0186654;  
OS Dictyostellium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.  
OX NCBI\_TaxID=44689;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=AX4;  
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,

RA Sucgang R., Barrilman M., Song J., Olsen R., Szafranski K., Xu Q.,  
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivoiro P.,  
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,

RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,  
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,

RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
RA Muzay D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,

RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
RA Loulsegue H., Mungall K., Oliver K., Price C., Quail M.A.,  
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,

RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,  
RA Shauly G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,

RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,  
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;  
RT "The genome of the social amoeba Dictyostellium discoideum";

RL Nature 0-0(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
KW Hypothetical protein.

SQ SEQUENCE 540 AA; 60799 MW; 0F48391CB55A35B5 CRC64;  
Query Match 15.7%; Score 102; DB 2; Length 540;  
Best Local Similarity 29.4%; Pred. No. 13;  
Matches 25; Conservative 18; Mismatches 38; Indels 4; Gaps 1;

QY 44 IKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESH-----RKEDLQREHSQKSDS 99  
Db 435 VRPVVKSSSKRKSEPTSEKSSKKKSKKSDVDVEMKEEPPVKEEKSKSKKS 494

QY 100 TKDVTATVLDKNNISSKSTNNPNK 124

Db 495 SKKEEPPVKEEKSSSKKEDKCKK 519

RESULT 19

Q9V7J0 DROME PRELIMINARY; PRT; 382 AA.  
AC Q9V7J0; O9GQ81;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE CG8421-PB, isoform B (Aspartyl beta-hydroxylase variant 2).

GN Name=Aeph; ORFNames=CG8421;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=2019606; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milahina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Testor C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";

RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Rettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;  
RA Dinichuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved  
RT isoform of Asph missing the catalytic domain share exons with  
RT junctin.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF003808; AAF58063.2; -; Genomic\_DNA.  
DR EMBL; AF289494; AGA0807.1; -; mRNA.  
DR Ensembl; CG8421; Drosophila melanogaster.  
DR FlyBase; FBgn0034075; Asph.  
DR FlyBase; FBgn0034075; CG8421.  
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03AEBFC6E8B CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 382;  
Best Local Similarity 24.5%; Pred. No. 9.7;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
Qy 1 EDPLPVYKGELEKGVQPDG-----ELSGPEGKKDAGYVI-----NLSKDTFIK 45  
Db 78 ELDLTPLESRRSK--VFDGWVDEHDEHGDHGVQPSGEALDDHDEHDDHDDHDEDEE 135  
Qy 46 PVFKIIEKKKEENKPT-----FVSVKKKNQPNVHNSHREKEDLQREHSQKSDS 99  
Db 136 PLTELELELEEEPTDEPAADDEVEEDEDENNA--GENYTAEDAEDEEEDND 193  
Qy 100 TKDVTATVLDKNNSKST 118  
Db 194 EGTVEATVEATTEATTEAT 212  
RESULT 20  
Q9V719 DROME PRELIMINARY; PRT; 556 AA.  
AC Q9V719;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG8421-PD, isoform D (CG8421-pe, isoform e).

GN Name=Asph; ORFNames=CG8421;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt E.G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Liao X., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Matt E., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon P., Nusaker D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Turner C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Rettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;  
RA Dinichuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved  
RT isoform of Asph missing the catalytic domain share exons with  
RT junctin.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF003808; AAF58063.2; -; Genomic\_DNA.  
DR EMBL; AF289494; AGA0807.1; -; mRNA.  
DR Ensembl; CG8421; Drosophila melanogaster.  
DR FlyBase; FBgn0034075; Asph.  
DR FlyBase; FBgn0034075; CG8421.  
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03AEBFC6E8B CRC64;



```
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirekas R., Smith E.,
RA Yu C., Rubin G.,
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289493; RAG0806.1; -; mRNA.
DR EMBL; AE003808; AAM70947.1; -; Genomic DNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; IEA.
DR GO; GO:0018193; P:peptidyl-amino acid modification; IEA.
DR InterPro; IPR007803; Asp Arg Hydrol.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001440; TPR-like helical.
DR InterPro; IPR001990; TPR-like helical.
DR Pfam; PF05118; Asp Arg Hydrol; 1.
DR PROSITE; PS0293; TPR REGION; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD5836F7F1 CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 785;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

Qy 1 EDFILPVYKGELEKGYQFDGW-----RISGEGKKDAGYVI-----NLKDTFIK 45
Db 78 EDLDTPLSESRFSK--VFDGWDEHRDHDGHDVQPSGEALDDHDDHDDHDEDEE 135
Qy 46 PVFKKIKKKEENKPT-----PDVSKKNQPNVNSHSHKEDLQREHSOKSDS 99
Db 136 PLTEELELEHEEEPTEDRPADEYEDDEENNA--GENITAEADREEREDND 193
Qy 100 TKDVTATVLDKNISKST 118
Db 194 EGTVEATVEATTEAT 212

RESULT 22
QHNR0 BACHK
ID Q6HNR0 BACHK PRELIMINARY; PRT; 954 AA.
AC Q6HNR0
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Possible internalin protein.
GN OrderedLocustNames=B79727_0463;
OS Bacillus thuringiensis (subsp. konkukian).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
OX NCBI_TaxID=180856;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.,
RT "Complete genome sequence of Bacillus thuringiensis 97-27.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017355; AAT63966.1; -; Genomic_DNA.
DR GO; GO:0009986; C:cell surface; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003885; LRR_cyst.
```

```
DR InterPro; IPR007092; LRR_SDS22.
DR InterPro; IPR006635; NEA_transpt.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00560; LRR_1; 8.
DR Pfam; PF05031; NEAT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00365; LRR_SD22; 8.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50978; NEAT; 1.
KW Complete proteome.
SQ SEQUENCE 954 AA; 108589 MW; 4F4CF8B44C9B355F CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 954;
Best Local Similarity 25.7%; Pred. No. 26;
Matches 43; Conservative 32; Mismatches 43; Indels 49; Gaps 9;

Qy 1 EDFILPVY--KGEL-----EKGYQF--DGWEIS-GFEG--- 28
Db 727 EIQVPPVYDLEGEIENIKLTSEDGTFNNGVIMSTPGKVKYKFDLSDSEISFNGTVI 786
Qy 29 -----KKDAGVINLSKDTPIKVPFKIEKKKEENKPTDVSKKDNPNVNSHSLN 80
Db 787 QNIVEKEEKEPTKEVEESKEEKEPT-KEVEESKEEKEPTKEVEESKEEKEPTKEVE 845
Qy 81 ESHRKEDLQREHSQKSDSTQVATVLDKN-----NISSKSTNNPN 123
Db 846 ES--KEEV--KEPTKEVEESKEEVAQEIEKSKBEINQSAFVQEQNVN 888

RESULT 23
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
EX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.B., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01001439; EAL42595.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;

Query Match 15.5%; Score 101; DB 2; Length 296;
Best Local Similarity 28.5%; Pred. No. 8;
Matches 35; Conservative 24; Mismatches 44; Indels 20; Gaps 5;

Qy 9 KGELEKGYQFDGWEISGFGKKDAGYVINLSKDTPIKVPFKIEKKE---ENKPTFDV 65
Db 145 EGDSEKKH-----DIPTNEGKENK----DTTKDKNDKEKKDVTNEGESSEKQKTEE 195
```





DR	GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR	GO; GO:0008270; F:zinc ion binding; IEA.
DR	GO; GO:0016567; P:protein ubiquitination; IEA.
DR	InterPro; IPR006209; EGF like.
DR	InterPro; IPR001841; Znf_RING.
DR	Pfam; PF00097; zf-C3HC4_1.
DR	SMART; SM00184; RING; 1.
DR	PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR	PROSITE; PS50089; ZF_RING_2; 1.
KW	Hypothetical protein; Metal-binding; Ub1 conjugation pathway; Zinc;
KW	Zinc-finger.
SQ	SEQUENCE 1130 AA; 131698 MW; FE6AAFE08C4CCDB6 CRC64;
 Query Match            15.4%; Score 100; DB 2; Length 1130; Best Local Similarity 26.1%; Pred. No. 41; Matches 30; Conservative 18; Mismatches 53; Indels 14; Gaps 3	
Qy	24 SGFGKKDAGV--INLSKOTFPVKFKIEEKEENKPTDVSKKDKNPQNHSQLNE 81
Dd	568 SFLSGSNDSYRIDRIYLPDPNVFYSSNNRNINDENVVEIVQEETRSDYNESHDTNE 627
Qy	82 SHRKEDLQREHSQS-----DSTKDVATVLDK-NISKSSTTNPNK 124
Dd	628 NNEKEBINREEGNEEKSRKRKTNDYDNKDVNNAAGSNNSHNHNNNNNK 682
 RESULT 27 Q90784.CHICK ID Q90784.CHICK PRELIMINARY; PRT; 1038 AA. AC Q90784; DT 01-NOV-1996 (TrEMBLrel. 01, Created) DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update) DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update) DE Claustrian. OS Gallus gallus (Chicken). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; OC Gallus. OX NCBI_TaxID=9031; RN [1] RP NUCLEOTIDE SEQUENCE. RC TISSUE=Brain; RX MEDLINE=94157526; PubMed=7906711; RA Burg M.A., Cole G.J.; RT "Claustrian, an antihedive neural keratan sulfate proteoglycan, is structurally related to MAP1B." RL J. Neurobiol. 25:1-22(1994). DR EMBL; X67778; CAA47988.1; -; mRNA. DR PIR; JC5497; JCS497. DR Ensembl; ENSGALG0000014999; Gallus gallus. SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;	
 Query Match            15.3%; Score 99.5; DB 2; Length 1038; Best Local Similarity 33.3%; Pred. No. 41; Matches 24; Conservative 20; Mismatches 19; Indels 9; Gaps 3	
Qy	38 LSKDTFTKPVFK-KIEEKEENKPTDVSKKDKNPQNHSQLNESHRKEDI-QRE----- 91
Dd	634 VTKEKSVKGVYAKPEEKQDEKPKGVSKKEKPLI---KGEKPKKEDIKGVKKEV 690
Qy	92 EHSQSDSTKDV 103
Dd	691 KEEKKAEKKEV 702
 RESULT 28 P91488.CABEL ID P91488.CABEL PRELIMINARY; PRT; 211 AA. AC P91488; DT 01-MAY-1997 (TrEMBLrel. 03, Created) DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update) DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) DE Hypothetical protein T33B3.5.	







```
Db 359 VIKLEERKQKEENSNNKFGSKRKKSKYNSESESDSRSDIDFDDHHNEHKSKI 418
Qy 94 -----SQKSDSTKDV-----ATVLDKNNISSKSTNNPN 123
Db 419 KNISKEKKNLSNEHSDNEKDVNRNKKIQTQLLDNNNNNNNNKNNNN 466

RESULT 35
ID O96229 PLAF7 PRELIMINARY; PRT; 951 AA.
AC O96229;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0680W.
GN Name=PFB0680W;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallem S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aton C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallem S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF001410; AAC71925.2; -; Genomic_DNA.
DR PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4P CRC64;

Query Match 15.1%; Score 98; DB 2; Length 951;
Best Local Similarity 27.1%; Pred. No. 48;
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

Qy 13 EKGQFDGWEI--SFEGKKDAGYVNLKDTFIKVPVKKEEKE-----EENK 60
Db 159 EKGQ-----DISNSNAENKCD-----VKEGVKELEEKKEEKESIDDHKVEENK 202
Qy 61 PTFFD----VSKKKDNPQVNHSQLNSHRKEDLQR-EEHSQKSDSKTQVATVLDKNNISS 115
Db 203 KSDDHKVEENKSDDHKVEENKSDDHKIEEVKKVEEHEDEE-----DKKEKS 253
Qy 116 KSTNNPNK 124
Db 254 ENKNDENK 262

RESULT 36
Q95S93 DROME
ID Q95S93 DROME PRELIMINARY; PRT; 556 AA.
AC Q95S93;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DB GM05229P.
GN Name=Aeph; ORFNames=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060905; AAL28453.1; -; mRNA.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 556 AA; 63089 MW; 95D32EAC57D11FE8 CRC64;

Query Match 15.0%; Score 97.5; DB 2; Length 556;
Best Local Similarity 24.5%; Pred. No. 29;
Matches 34; Conservative 28; Mismatches 52; Indels 25; Gaps 5;

Qy 1 EDFLPVYKGELEKGYQFGW-----EISGFSKKDAGYVI-----NLSKDTFIK 45
Db 78 EDLDTPLSESRFSK--VFDGWVDEHDEHGDHGVQEPSEALDDHDDHDDHDEDEE 135
Qy 46 PVFKKIEEKEENKPT-----FDVSKKKDNPQVNHSQLNSHRKEDLQREHSQKSDS 99
Db 136 PLTELEEELEEEEPTEDEPAADVEYDEDEENNA--GENITAEADAVEEEEDND 193
Qy 100 TKQVATVLDKNNISSKT 118
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 37
YH00 YEAST
ID YH00 YEAST STANDARD; PRT; 1345 AA.
AC P38800;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical 149.7 kDa protein in IRE1-KSPI intergenic region.
GN OrderedLocusNames=YHR080C;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latrelle P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaekis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RL VIII.";
RL Science 265:2077-2082(1994).
CC -!- SIMILARITY: Contains 1 GRAM domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
```

```

CC EMBL; U10556; AAC6895.1; -, Genomic_DNA.
DR PIR; S46817;
DR Germline; 139397;
DR Ensembl; YHR080C; Saccharomyces cerevisiae.
DR SCD; S000001122; YHR080C.
DR GO; GO:0005739; C.mitoochondrion; IDA.
DR InterPro; IPR004182; GRAM.
DR SMART; PF02893; GRAM; 1.
DR Pfam; SM00568; GRAM; 1.
KW Complete proteome; Hypothetical protein; Transmembrane.
FT TRANSMEM 1198 1218 Potential.
FT DOMAIN 548 617 GRAM.
FT COMBIAS 679 711 Asp-rich.
SQ SEQUENCE 1345 AA; 149680 MW; 2FDAB94A686564C2 CRC64;

Query Match 15.0%; Score 97.5; DB 1; Length 1345;
Best Local Similarity 27.0%; Pred. No. 76;
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 9 KGELEKGYQDCWEISGFGKK-DAGYVINLSKDTFKIPVFKIEEKEENKPTFDVSK 67
D 1109 KGAIEKG-----SVGEQKVSVDYMLSELRDII-----SRAKSKPKVYKWK 1149
QY 68 KDNQVNHSLNESHKEDLQREHSQKSDTKDVTATVLD--KNNISSTNN--NP 122
D 1150 SHDKRPHSKVE-----QKSESRKSDDNKDILTHILDVQVNFSSRIFNKKLLSP 1201
QY 123 NK 124
D 1202 QK 1203

RESULT 38
Q427T5 PLABE
ID Q427T5 PLABE PRELIMINARY; PRT; 895 AA.
AC Q427T5;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DE Hypothetical protein (Fragment).
GN ORFNames=PB000019.00.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAI01000607; CAH94732.1; -, Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1028 AA; 120493 MW; 866A9FFFC427612 CRC64;

Query Match 14.9%; Score 97; DB 2; Length 1028;
Best Local Similarity 25.6%; Pred. No. 62;
Matches 40; Conservative 20; Mismatches 48; Indels 48; Gaps 6;

QY 8 YKGELEKGYQDCWEISGFGKKDAGYVINLSKDTFKIPVFK-----LEKKKE 56
D 574 YSEYKGEK-----KKPVFEGKK-----NLKKKKKKKLNLDKGIKNNKINIEITVKEENS 624
QY 57 ENKPTFDVSKKDNQVNHSLNESHKEDLQREHSQKSDTKDVTATVLD-----EHSQKSDS----- 99
D 625 EKKKEAYDETNNKKIKSKNDLVPKKIEEKEENKPTFDVSK-----LEKKKEENKSNIKK 684
QY 100 ----TKDVTATVLDKNNI-----SKSTNNPNK 124
D 685 YKLFKDFEHLKEDNIEIGKKNESSTYKKNDSNK 720

RESULT 40
Q6BXEI DEBHA
ID Q6BXEI DEBHA PRELIMINARY; PRT; 905 AA.
AC Q6BXEI;
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DE Similar to tr|Q8TG35 Candida albicans Mnn4p.
GN OrderedLocusNames=DEHA0B03718g;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
EX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., March C., Neugeglise C., Talla E.,
RA Goffard N., Prangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,

```

```

QY 101 KDVTATVLDKNNISKSTNNPNK 124
D 630 KKVKNI-----KENIFLDVKNKQTNK 650

RESULT 39
Q424Q1 PLABE
ID Q424Q1 PLABE PRELIMINARY; PRT; 1028 AA.
AC Q424Q1;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=PB0000814.00.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAI01000607; CAH94732.1; -, Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1028 AA; 120493 MW; 866A9FFFC427612 CRC64;

Query Match 14.9%; Score 97; DB 2; Length 1028;
Best Local Similarity 25.6%; Pred. No. 62;
Matches 40; Conservative 20; Mismatches 48; Indels 48; Gaps 6;

QY 8 YKGELEKGYQDCWEISGFGKKDAGYVINLSKDTFKIPVFK-----LEKKKE 56
D 574 YSEYKGEK-----KKPVFEGKK-----NLKKKKKKKLNLDKGIKNNKINIEITVKEENS 624
QY 57 ENKPTFDVSKKDNQVNHSLNESHKEDLQREHSQKSDTKDVTATVLD-----EHSQKSDS----- 99
D 625 EKKKEAYDETNNKKIKSKNDLVPKKIEEKEENKPTFDVSK-----LEKKKEENKSNIKK 684
QY 100 ----TKDVTATVLDKNNI-----SKSTNNPNK 124
D 685 YKLFKDFEHLKEDNIEIGKKNESSTYKKNDSNK 720

RESULT 40
Q6BXEI DEBHA
ID Q6BXEI DEBHA PRELIMINARY; PRT; 905 AA.
AC Q6BXEI;
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DE Similar to tr|Q8TG35 Candida albicans Mnn4p.
GN OrderedLocusNames=DEHA0B03718g;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
EX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., March C., Neugeglise C., Talla E.,
RA Goffard N., Prangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,

```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 13.038 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773  
Perfect score: 651  
Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues 283416  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	2140	2 F95074	serine proteinase,
2	648	99.5	2144	2 A97942	metalloproteinase,
3	108	16.6	558	2 T18467	hypothetical prote
4	93.5	15.3	1038	2 J54977	claustrin - chicke
5	99	15.2	211	2 T25911	hypothetical prote
6	98	15.1	665	2 B71609	hypothetical prote
7	97.5	15.0	1345	2 S46817	hypothetical prote
8	95	14.6	348	2 I37271	cyclicin II - human
9	94	14.4	210	2 T28771	hypothetical prote
10	93	14.3	535	2 T37189	hypothetical prote
11	92.5	14.2	219	2 B72291	hypothetical prote
12	92.5	14.2	312	2 G81339	probable membrane
13	91	14.0	253	2 T32879	hypothetical prote
14	90	13.8	325	2 T18283	hypothetical prote
15	89.5	13.7	508	2 B81594	hypothetical prote
16	89.5	13.7	508	2 E86549	hypothetical prote
17	89.5	13.7	508	2 C72074	hypothetical prote
18	89	13.7	528	2 E96795	unknown protein F2
19	89	13.7	622	2 A90570	lipoprotein (impor
20	89	13.7	1397	2 T10466	DNA topoisomerase
21	88.5	13.6	301	2 T33068	hypothetical prote
22	88.5	13.6	385	2 T20410	hypothetical prote
23	88	13.5	276	2 T23451	hypothetical prote
24	88	13.5	700	2 S67610	probable membrane
25	87.5	13.4	644	2 T47835	hypothetical prote
26	87.5	13.4	1702	2 A41859	Iga-specific metal
27	87.5	13.4	2523	2 T18477	hypothetical prote
28	87	13.4	635	2 F71621	hypothetical prote
29	87	13.4	2464	1 QRMSP1	microtubule-associ

30	86.5	13.3	762	2 G88436	protein T04A8.13 [
31	86.5	13.3	791	2 T24435	hypothetical prote
32	86.5	13.3	2485	1 H71621	serine/threonine-s
33	86	13.2	1016	2 T19006	ankyrin related pr
34	85	13.1	335	2 T33457	hypothetical prote
35	85	13.1	1166	2 H71609	hypothetical prote
36	85	13.1	1202	1 S05362	probable DNA-direc
37	85	13.1	1976	2 T56555	sodium channel pro
38	85	13.1	3006	2 T28625	variant-specific s
39	84.5	13.0	264	2 A44969	circumsporozoite p
40	84.5	13.0	645	2 E89883	conserved hypothet
41	84.5	13.0	1694	2 H64106	Iga-specific metal
42	84	12.9	312	2 T25994	hypothetical prote
43	84	12.9	417	2 T49847	hypothetical prote
44	83.5	12.8	143	2 T29740	hypothetical prote
45	83.5	12.8	281	2 T29150	hypothetical prote

ALIGNMENTS

RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heuvelink, A.E.; Hickey, E.K.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, M.; J.D.; Mayam, L.A.; White, O.; Salzberg, S.L.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%;	Score	651;	DB	2;	Length	2140;
Best Local Similarity	100.0%;	Pred. No.	7.5e-44;				
Matches	124;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFEKGDAGTVINLSKDTPIKPVFKLEEKKSEENK	60				
Db	1983	EDFILPVYKGELEKGYQFDGWEISGFEKGDAGTVINLSKDTPIKPVFKLEEKKSEENK	2042				
Qy	61	PTFDVSKKDNQVNHSQLNESHKEDLQREHSHKSDSTKVATVLDKNNISKSTTN	120				
Db	2043	PTFDVSKKDNQVNHSQLNESHKEDLQREHSHKSDSTKVATVLDKNNISKSTTN	2102				
Qy	121	NPVK 124					
Db	2103	NPVK 2106					

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burtett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234







C:Genetics:  
A:Gene: CESP:C02H7.1  
A:Map position: X  
A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 14.3%; Score 93; DB 2; Length 535;  
Best Local Similarity 24.2%; Pred. No. 4.9;  
Matches 30; Conservative 22; Mismatches 52; Indels 20; Gaps 3;

Qy 11 ELEKGYQPDGWEISGPEGKDGAGYVINLSKDTPIKVPFKKIEEKEENKPTFDVSKKD 70  
Db 97 ETNKMQLMGTNATNSFRNGTG-----BEKKKKVKEDKKGDEBEKST---TKKRS 146  
Qy 71 NPQVNSQLNSHREKDLQREHSQKSDSTK-----DVTATVLDKNNISSKSTTN 120  
Db 147 SKKETHEKEKSEKKSAEKKKKSSSSSKERHKSSDRSSEKSKSEKKKKEKSTTD 206  
Qy 121 NPNK 124  
Db 207 EKPK 210

RESULT 11  
B72291  
hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: UNIPROT:Q9X0M6; UNIPARC:UPI00000C12ED; GB:AE001771; GB:AE000512; NID  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1142

Query Match 14.2%; Score 92.5; DB 2; Length 219;  
Best Local Similarity 28.1%; Pred. No. 2;  
Matches 36; Conservative 24; Mismatches 35; Indels 33; Gaps 8;

Qy 10 GELEKGYQ--PDGWEISG-----PEGKDGAGYVIN-LSKDTPIKVPFKKIEEKEENK 60  
Db 89 GEESVSRYDLFAGFGVRGTPPTFPFKGEGGLGYPGVKDNFIK-ILKYVAQLKED-- 145  
Qy 61 PTFDVSKKDNPQVNSQLNSHREK-----LQREHSQKSDS-----TKD 102  
Db 146 --FOTLYKKDDPFVGEPLIIEIF-KEDADFVLEKDNVAKVDTPVNEVRDRIRYVTDSPD 202  
Qy 103 VTATVLDK 110  
Db 203 VAKTLQEK 210

RESULT 12  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: G81339  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hye  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: G81339  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-312 <PAR>  
A:Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL139076; GB:AL111168; NID  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0692c

Query Match 14.2%; Score 92.5; DB 2; Length 312;  
Best Local Similarity 27.6%; Pred. No. 2.9;  
Matches 35; Conservative 26; Mismatches 49; Indels 17; Gaps 6;

Qy 11 ELEKGYQPDGWEISGPEGKDGAGYVINL---SKDTPIKVPFKKIEEKEENKPT---FD 64  
Db 59 ELDEVEYSKUTKSKSNLYLKED---LINVLEEKQSLAKKIFSKMKERRKEENKTKKNFL 115  
Qy 65 VSKKKDNP-----QVNSQLNSHREKDLQREHSQKSDSTKDV--TATVLDKNNISSK-- 116  
Db 116 FSRKKAENEIKNIQTQIQTKSNQATTQTQKKEKELTNSIEKIQKTETKIQKPLIEKKL 175  
Qy 117 STTNPN 123  
Db 176 DVKNQPN 182

RESULT 13  
T32879  
hypothetical protein C17F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T32879  
R:Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C17F3.  
A:Reference number: Z21240  
A:Accession: T32879  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-253 <GAT>  
A:Cross-references: UNIPROT:O44948; UNIPARC:UPI0000074BB9; EMBL:AF043692; PIDN:AAB97531.  
A:Experimental source: strain Bristol N2; Clone C17F3  
C:Genetics:  
A:Gene: CESP:C17F3.3  
A:Map position: 1  
A:Introns: 41/1

Query Match 14.0%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 3;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

Qy 46 PVPKKIEKKK---EENKPTFDVSKK-----KDNQVNSQLNSHSHR----- 84  
Db 61 PVAKVKEKKKEEKEEKKADDEKKTEEKDDKKSKTEEKDKISVKKYQTKSERKDKK 120  
Qy 85 ---KEDLQREHSQKSDSTKDVATVATVLDKNNISSK 116  
Db 121 DEKDEKESKESEKSKDEEKKKDEKDEKDEKK 155

RESULT 14  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18283  
R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugu  
Genetics 148, 1117-1125, 1998  
A:Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A:Reference number: Z14684; MUID:98198836; PMID:9539429  
A:Accession: T18283  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-325 <RIE>  
A:Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; P1





QY 109 DK 110  
DB 370 EK 371

## RESULT 23

T23451  
hypothetical protein K08E3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23451  
R:McMurray, A.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23451  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-276 <WIL>  
A:Cross-references: UNIPROT:Q9XUT1; UNIPARC:UPI0000077D88; EMBL:Z81568; PIDN:CAB04590.1;  
A:Experimental source: clone K08E3  
C:Genetics:  
A:Gene: CESP:K08E3.2  
A:Map position: 3  
A:Introns: 81/3; 102/3; 169/1; 211/2

Query Match 13.5%; Score 88; DB 2; Length 276;  
Best Local Similarity 31.5%; Pred. No. 5.8;  
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 3;

QY 38 LSKDTFIRPVFKKIEKKKEE--NKPTDVSKKNDPQVNHSQLSHKEDLQREHSQ 95  
DB 60 LEKSDVEKKKGEKKEEKKGEKKEDKKEKKD-----SKEDKKDEDHEK 107

QY 96 KSDSTKDVATVLDKNNISSKSTNNPNK 124  
DB 108 KK-----TAEKENNEKKDKNK 128

## RESULT 24

S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D2483  
C:Species: Saccharomyces cerevisiae  
C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: S67610  
R:Wambutt, R.; Wedler, H.; Wedler, B.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67608  
A:Accession: S67610  
A:Molecule type: DNA  
A:Residues: 1-700 <WAM>  
A:Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069BFF; EMBL:Z74122; NID:G1431087; PIDN:G1431087; PIDN:G1431087  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:BREL; MIPS:YDL074c  
A:Cross-references: SGD:S0002232  
A:Map position: 4L  
C:Keywords: transmembrane protein  
F:69-85/Domain: transmembrane #status predicted <TMM>

Query Match 13.5%; Score 88; DB 2; Length 700;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 31; Conservative 24; Mismatches 40; Indels 36; Gaps 5;

QY 11 ELEKGYQDGEWISGFEKK-----DAGVIVNL-SKDTF---IK 45  
DB 447 DLEKGF-----ELSDLTHKYSBIINHSVISKLTVEKTKADQKYFAAMRSKDSILIBIK 502

QY 46 PVFKKIEKKKEENKPTDVSKKNDPQVNHSQLSHKEDLQREHSQSDSKDVTA 105  
DB 503 TLSKSLSKNEL-----ILQKSDRLLOKQIGNLHKQLDLSQNNERLIDSKTYTL 555  
QY 106 TVLDKNNISSK 116

DB 556 KIIDLNTSTK 566

## RESULT 25

T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Eartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NVA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL1138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 13.4%; Score 87.5; DB 2; Length 644;  
Best Local Similarity 24.4%; Pred. No. 16;  
Matches 31; Conservative 18; Mismatches 51; Indels 27; Gaps 3;

QY 2 DFILPVVKGLEKGYQDGEWISGFEKKDAGYVIVNLKDTFIKPVFKKIEKKKEENKP 61  
DB 537 DFLKRIKKNSPQK-----ETTSKNQKNDGNV-----KKENDHOK 573

QY 62 TFDVSKKNDPQVNHSQLSHKEDLQREHSQKSDS-----TKDVTATVLDKNNISSK 117  
DB 574 KSDGNVKKENSKVPRELRSSTGKKVVENNSKSSKKRQTETAEVATGKRGESGK 633

QY 118 TTNPNK 124  
DB 634 DDQPRK 640

## RESULT 26

A41859  
IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae  
C:Species: Haemophilus influenzae  
A:Variety: strain HK715  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41859  
R:Poulsen, K.; Reinholdt, J.; Killian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae  
A:Reference number: A41859; MUID:92234949; PMID:1373717  
A:Accession: A41859  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1702 <POU>  
A:Cross-references: UNIPROT:P45384; UNIPARC:UPI000012D3F0; GB:M87489; NID:G148906; PIDN:G148906  
A:Experimental source: strain HK715  
A:Note: sequence extracted from NCBI backbone (NCBIP:97282)  
C:Superfamily: IGA-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 13.4%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 49;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 36 INLSKDTFIRPVFKKIEKKKEENKPTFDVSKKNDPQVNHSQLSHKEDLQREHS- 94  
DB 1296 INTGSATAITETAESKDKPTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 95 --QKSDSTKDVATVLDKNNISSKSTNNPNK 124

DB 1356 SQPQTSABETTAASTDETTIADNSKSKPNR 1387



QY 72 PQVNHSQLNESHKEDLQREHSHQSDSTKVDTATVLDKNNISKSTNNPK 124  
DB 84 EKEDGHEKEDKEDKENENDEKKEKSKDDKESKEDKKEKTKTNECK 136

## RESULT 31

T24435

hypothetical protein T04A8.13 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T24435

R:Palmer, S.

submitted to the EMBL Data Library, August 1994

A:Reference number: Z19889

A:Accession: T24435

A&gt;Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-791 &lt;WIL&gt;

A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PIDN:CAA84732.2;

A:Experimental source: clone T04A8

C:Genetics:

A:Gene: CESP:T04A8.13

A:Map position: 3

A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

## Query Match

Best Local Similarity 13.3%; Score 86.5; DB 2; Length 791;

Matches 29; Conservative 22; Mismatches 59; Indels 3; Gaps 2;

Query Match 13.3%; Score 86.5; DB 2; Length 791;

Matches 29; Conservative 22; Mismatches 59; Indels 3; Gaps 2;

QY 13 EKGYPDQWEISGPEK-KDAGYVNLKDTFTKPKVFKIEKKEENKPTFDVSKKON 71

DB 26 KEGLGMDQKEIVGDDKKDKBAKRERKLQDEPAB--LKKDESKDEBAKKEKKEKKEE 83

QY 72 PQVNHSQLNESHKEDLQREHSHQSDSTKVDTATVLDKNNISKSTNNPK 124

DB 84 EKEDGHEKEDKEDKENENDEKKEKSKDDKESKEDKKEKTKTNECK 136

## RESULT 32

H71621

serine/threonine-specific protein kinase (EC 2.7.1.1-) PPB0150c - malaria parasite (Plasm

C:Species: Plasmodium falciparum

C&gt;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004

C:Accession: H71621

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.

; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: H71621

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2485 &lt;GAR&gt;

A:Cross-references: UNIPROT:O96134; UNIPARC:UPI0000172553; GB:AE001376; GB:AE001362; NID

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0150c

C:Superfamily: protein kinase homology

C:Keywords: phosphotransferase

F:2087-2352/Domain: protein kinase homology &lt;KIN&gt;

## Query Match

Best Local Similarity 13.3%; Score 86.5; DB 1; Length 2485;

Matches 36; Conservative 27; Mismatches 31; Indels 57; Gaps 9;

Query Match 13.3%; Score 86.5; DB 1; Length 2485;

Matches 36; Conservative 27; Mismatches 31; Indels 57; Gaps 9;

QY 11 ELEKGYPDQW-----RISGFE-----GKKDAGYVNLKDTFTKPKVFKIEE 53

DB 1977 ENERGDQKGYEEMNGDKNEEMNGDKNEEMNGDKNGG--IN-----EE 2020

QY 54 KK-----EENKPTFDVSKKONPVN-----HSQLNESHKEDLQ-----REH----- 93

DB 2021 HKNEGINEEHKDEL--INKEHKQVERINEEHKQVERINEEHKQVERINEEHKQVER 2079

QY 94 ---SQKSDSTKVDTATVLDKNNISKSTNN 121  
DB 2080 INEEHKNKGNKLTVHNNKNNISNNYND 2110

## RESULT 33

T19006

ankyrin related protein C06C3.1 - Caenorhabditis elegans

C:Species: myosin-light-chain-phosphatase (EC 3.1.1.353)

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T19006; T22086

R:Berks, M.

submitted to the EMBL Data Library, August 1994

A:Reference number: Z19058

A:Accession: T19006

A&gt;Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-1016 &lt;WIL&gt;

A:Cross-references: UNIPROT:Q17718; UNIPARC:UPI0000081D09; EMBL:Z36719; PIDN:CAA85318.1

A:Experimental source: clone C06C3

R:Matthews, P.

submitted to the EMBL Data Library, January 1995

A:Reference number: Z19510

A:Accession: T22086

A&gt;Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-1016 &lt;WIL&gt;

A:Cross-references: UNIPARC:UPI0000081D09; EMBL:Z47809; PIDN:CAA87782.1; GSPDB:GN000020

A:Experimental source: clone F42A8

C:Genetics:

A:Gene: CESP:C06C3.1

A:Map position: 2

A:Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 569/2; 718/3; 766/3; 833/3;

C:Keywords: phosphoric monoester hydrolase

## Query Match

Best Local Similarity 13.2%; Score 86; DB 2; Length 1016;

Matches 32; Conservative 30; Mismatches 35; Indels 34; Gaps 8;

Query Match 13.2%; Score 86; DB 2; Length 1016;

Matches 32; Conservative 30; Mismatches 35; Indels 34; Gaps 8;

QY 12 LEKGYPQ-----PDGWEISGPEK-KDAGYVNLKDTFTKPKVFKIEKKEE---NKPTF 63

DB 267 LENGAEISDLTFTGADVLGVADKCIDLYVELA-DTV-----KVQNKRKSPGSGSQPT 319

QY 64 DVSKKONPVNHQO-LNESHKEDLQ-REHSHOK-----SDSTKVDTA 105

DB 320 SILGEKMRMPSHEHVLTSEKRDQLQKHQSERNEFLHSPSTASVGSSTSSNTNTT 379

QY 106 T-VLDKNNISS 115

DB 380 TIVIGENDISA 390

## RESULT 34

T33457

hypothetical protein F36H12.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T33457

R:Blanchard, M.; Bradshaw, H.; Stellyes, L.

submitted to the EMBL Data Library, July 1998

A:Description: The sequence of C. elegans cosmid F36H12.

A:Reference number: Z21346

A:Accession: T33457

A&gt;Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-335 &lt;BLA&gt;

A:Cross-references: UNIPROT:O76719; UNIPARC:UPI0000076COA; EMBL:AF078790; PIDN:AAC26930

A:Experimental source: strain Bristol N2; clone F36H12

C:Genetics:

A:Gene: CESP:F36H12.3

A:Map position: 4





Job time : 16.038 secs

```
Db 2194 RKTFGHTYCKPSKLTVKCKNNHCDNSKPNDCRNINSISAEDIEKRSNSTQDVTMSVS 2253
QY 109 DKNNTSSK 116
Db 2254 DSNTNGNK 2261

RESULT 39
A44969
circumsporozoite protein precursor - Plasmodium yoelii nigeriensis
C:Species: Plasmodium yoelii nigeriensis
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-1995
C:Accession: A44969
R:Colomer-Gould, V.; Enea, V.
Mol. Biochem. Parasitol. 43, 51-58, 1990
A:Title: Plasmodium yoelii nigeriensis circumsporozoite gene structure and its implicati
A:Reference number: A44969; MUID:91148645; PMID:2290446
A:Accession: A44969
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <COL>
A:Cross-references: UNIPARC:UPI000177F73; GB:M32350
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
F:190-242/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 13.0%; Score 84.5; DB 2; Length 264;
Best Local Similarity 28.2%; Pred. No. 10;
Matches 29; Conservative 14; Mismatches 29; Indels 31; Gaps 5;
QY 45 KVPFKK-----IEKKK-----EENKPTFDVSKKDNQVNHSQLNESHKEDLQRE----- 92
Db 75 KPAEKKDLLPKKEKKDDLPKEKKDDLP-----KEKKDDLPKEKKDA 127
QY 93 -----HSQKSDSKDVTATVL--DKNNISSKSTNNPN 123
Db 128 PFGNKDAPKEKKADPPKEAQNKLKQPDNNNNNNNNNNNN 170

RESULT 40
E89883
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E89883
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: E89883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-645 <KUR>
A:Cross-references: UNIPROT:Q99UK5; UNIPARC:UPI00000CAE3; GB:BA000018; PID:g13700929; E
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0976

Query Match 13.0%; Score 84.5; DB 2; Length 645;
Best Local Similarity 28.6%; Pred. No. 28;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;
QY 31 DAGYVNL-SKDTFLKVPFKIEKKKEENKPTDV-----SKKDNQVNHSQLNESH 84
Db 441 DQYHVRIVDKAFTKANTDKSKKEQQDNKAKKATPATPSKPTSPVEKESQKDSOK 500
QY 85 KEDLQ-----REHSQKSDSKDVT--ATVLDKNNISSKSTNNPNK 124
Db 501 DNNKQLPSVEKENDASSESGKDKTPATKPTGEVSSSTT--PTK 543

Search completed: April 24, 2006, 15:01:11
```

**THIS PAGE BLANK (user)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 9.70917 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISKSTNNPNK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /SIDSS/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /SIDSS/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /SIDSS/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /SIDSS/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /SIDSS/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /SIDSS/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /SIDSS/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /SIDSS/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	16.3	354	7 US-11-189-817-2	Sequence 2, Appli
2	101	15.5	651	7 US-11-128-660-1	Sequence 1, Appli
3	86	13.2	443	7 US-11-188-298-1015	Sequence 1015, Ap
4	85	13.1	472	6 US-10-793-626-658	Sequence 658, App
5	85	13.1	720	6 US-10-793-626-2058	Sequence 2058, Ap
6	85	13.1	746	6 US-10-793-626-652	Sequence 652, App
7	84.5	13.0	501	6 US-10-485-517-381	Sequence 381, App
8	84.5	13.0	645	6 US-10-485-517-244	Sequence 244, App
9	84.5	13.0	1694	7 US-11-052-554A-83	Sequence 83, Appl
10	82	12.6	439	7 US-11-188-298-15964	Sequence 15964, A
11	82	12.6	439	7 US-11-188-298-16606	Sequence 16606, A
12	82	12.6	886	7 US-11-087-099-11456	Sequence 11456, A
13	79.5	12.2	140	7 US-11-096-568A-4771	Sequence 4771, Ap
14	79.5	12.2	943	6 US-10-475-204-34	Sequence 34, Appl
15	79.5	12.2	1036	7 US-11-096-568A-28315	Sequence 28315, A
16	79.5	12.2	1070	7 US-11-096-568A-28314	Sequence 28314, A
17	79.5	12.2	1276	7 US-11-096-568A-28313	Sequence 28313, A
18	78.5	12.1	258	7 US-11-098-686-10475	Sequence 10475, A
19	78.5	12.1	700	7 US-11-196-475-74	Sequence 74, Appl
20	78.5	12.1	1155	6 US-10-793-626-1780	Sequence 1780, Ap
21	77.5	11.9	405	6 US-10-485-517-239	Sequence 239, App
22	77.5	11.9	627	6 US-10-873-528-191	Sequence 191, App
23	77	11.8	663	7 US-11-196-475-70	Sequence 70, Appl
24	77	11.8	663	7 US-11-196-475-78	Sequence 78, Appl
25	76.5	11.8	439	7 US-11-188-298-17915	Sequence 17915, A

ALIGNMENTS

RESULT 1

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US2006030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM  
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES  
; FILE REFERENCE: 275601USO  
; CURRENT APPLICATION NUMBER: US/11/189,817  
; CURRENT FILING DATE: 2005-07-27  
; PRIOR APPLICATION NUMBER: 60/598,062  
; PRIOR FILING DATE: 2004-08-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-11-189-817-2

Query Match 16.3%; Score 106; DB 7; Length 354;  
Best Local Similarity 23.5%; Pred. No. 0.046;  
Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;

Qy 7 VYKGELEKGYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKIEKKKEE- 57  
Db 181 VLKAKRASSYDIILGWEPFGGVPPEKKKENMLSHLYVSSKDKENISKENDVDLDEKEEA 240  
Qy 58 -----ENKPTFDVSKKDNPNQVNHSQLNESHREKEDLQREHSQKSDSTKDVTA 105  
Db 241 ETEEEELKKNVEETSEISEDEEBEKEEENDKKKEQKESNNNDQKDMEA 300  
Qy 106 TVLDKNNISKSSTNN 121  
Db 301 -----QNLISKNNNN 311

RESULT 2

US-11-128-660-1  
; Sequence 1, Application US/11128660  
; Publication No. US2006002432A1  
; GENERAL INFORMATION:  
; APPLICANT: Statens Serum Institut  
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmodium falciparum  
; TITLE OF INVENTION: falciparum  
; FILE REFERENCE: 15007dk

	INTERLUCE	257	CONSERVATIVE	207	PLURALITIES	337	INDICES	201	GAPS
Qy	20	GWEISGF--EGKQAG-----	YVNL	SKDTPIKVPFKKIEBKKEENKCTFDVSKKQNP	72				
Db	493	GWEFGGVPHEHKCBENMLSHLYVSSKKNISKENDDLVDE--KEEAAEETEEEELEEKNE	551						
Qy	73	QVNHSGLN-----	ESHKED	LOREHHSQKSDSTKDVTVATVLDKNNISKSTT	119				
Db	552	EETESISDEDEEEEEEKEENKKEQKEQENNNDDQKWEA-----	QNLISKQNN	606					
Qy	120	NN	121						
Db	607	NN	608						

	matches	22,	conservative	23,	mis-matches	10,	inserts	12,	gaps
Qy	8	YKBELEKGYPQGWISGIFSGEKDAGVYNLSKDTFI-----	44						
		:                 : : :							
Dd	42	YBEAVEDVGSPFGSSIPFEGIEDSLIFKADPPSYAEI PWEGIGRVGYIYKDEPYOA	101						
Qy	45	--KPWFKEIEKKKEEN-----KPTFDVSKKD-----	70						
		:   :   :   :   :							
Dd	102	DPFGILKRVLRELEKEGLKAHIGPEPEYVIFKNGTWELHIPDSGGVPDLVGLDKAREIR	161						
Qy	71	-----NPQWNSQLNESHREDLQREHSOKSD---STKDVATATVLD	109						
Dd	162	REIALYMPYLGIKPELVHEHYEYGAOHBIDFYDEALRTADNIYSFKHVVKAAVE	215						

RESULT 6  
US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258







; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: PROTEIN INTERACTIONS IN VERTEBRATE CELLS  
; FILE REFERENCE: HMV-056.25  
; CURRENT APPLICATION NUMBER: US/10/475,204  
; CURRENT FILING DATE: 2003-10-17  
; PRIOR APPLICATION NUMBER: PCT/US02/13008  
; PRIOR FILING DATE: 2002-09-27  
; PRIOR APPLICATION NUMBER: 60/285,509  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-475-204-34

Query Match 12.2%; Score 79.5; DB 6; Length 943;  
Best Local Similarity 23.1%; Pred. No. 30;  
Matches 36; Conservative 16; Mismatches 55; Indels 49; Gaps 5;  
  
QY 6 PVYGELEKGYQFDGWEISFGKGDAGVYVNLSDTKPIKVPFKIIEKK-----EEN 59  
Db 375 PSDKTVLDSYALIDETVNNYSTKYMYSKNAEFSKRTIKQKQKFWAKPAEQ- 433  
QY 60 KPTFDVSKKD-NPQVNHSQLNESHKEDLQREH-----SQKSDSTK 101  
Db 434 ---LDVGQSKDENIHTSHITQDEFQKSDRNMEHEMGNDVCVQKMPFVGSKKSSTK 490  
QY 102 D-----VTATVLDKNNISSK 116  
Db 491 DKESKKRPFSESKNKLVPVEVTSVTKSRISR 526

RESULT 15  
US-11-096-568A-28315  
; Sequence 28315, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28315  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1036)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712010  
US-11-096-568A-28315

Query Match 12.2%; Score 79.5; DB 7; Length 1036;  
Best Local Similarity 23.2%; Pred. No. 33;  
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;  
  
QY 27 EGKDGAGVYVNLSDTKP-----IKPVFKIIEKKKEENKPTFDVSKKDNPOVNHSQLN 80  
Db 552 DGDSDKKVMEVGKSSDGSVEMKPTAELSLEDVDENASKTVDVVKQETGSPDTKKKEGA 611  
QY 81 ESHRKEDLQREH-----SQKSDSTKDVATVLDKNNISSK 116  
Db 612 SSSSKKDTKTGDKKAEKKNSETMSEGGKIDRNTDEK 650

RESULT 16  
US-11-096-568A-28314  
; Sequence 28314, Application US/11096568A  
; Publication No. US20060048240A1

; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28314  
; LENGTH: 1070  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1070)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712009  
US-11-096-568A-28314

Query Match 12.2%; Score 79.5; DB 7; Length 1070;  
Best Local Similarity 23.2%; Pred. No. 35;  
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;  
  
QY 27 EGKDGAGVYVNLSDTKP-----IKPVFKIIEKKKEENKPTFDVSKKDNPOVNHSQLN 80  
Db 586 DGDSDKKVMEVGKSSDGSVEMKPTAELSLEDVDENASKTVDVVKQETGSPDTKKKEGA 645  
QY 81 ESHRKEDLQREH-----SQKSDSTKDVATVLDKNNISSK 116  
Db 646 SSSSKKDTKTGDKKAEKKNSETMSEGGKIDRNTDEK 684

RESULT 17  
US-11-096-568A-28313  
; Sequence 28313, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28313  
; LENGTH: 1276  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1276)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712008  
US-11-096-568A-28313

Query Match 12.2%; Score 79.5; DB 7; Length 1276;  
Best Local Similarity 23.2%; Pred. No. 43;  
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;  
  
QY 27 EGKDGAGVYVNLSDTKP-----IKPVFKIIEKKKEENKPTFDVSKKDNPOVNHSQLN 80  
Db 792 DGDSDKKVMEVGKSSDGSVEMKPTAELSLEDVDENASKTVDVVKQETGSPDTKKKEGA 851  
QY 81 ESHRKEDLQREH-----SQKSDSTKDVATVLDKNNISSK 116  
Db 852 SSSSKKDTKTGDKKAEKKNSETMSEGGKIDRNTDEK 890

RESULT 18  
US-11-098-686-10475  
; Sequence 10475, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES

```
/ TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
/ FILE REFERENCE: 09531-128001
/ CURRENT APPLICATION NUMBER: US/11/098,686
/ CURRENT FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: PCT/US03/31318
/ PRIOR FILING DATE: 2003-10-01
/ PRIOR APPLICATION NUMBER: US 60/416,395
/ PRIOR FILING DATE: 2002-10-04
/ NUMBER OF SEQ ID NOS: 11433
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10475
/ LENGTH: 258
/ TYPE: PRT
/ ORGANISM: Lawsonia intracellularis
US-11-098-686-10475

Query Match          12.1%; Score 78.5; DB 7; Length 258;
Best Local Similarity 22.0%; Pred. No. 8.1;
Matches 20; Conservative 26; Mismatches 30; Indels 15; Gaps 3;

QY 38 LSKDFTFKPVFKIEEKEENKPT-FDVSKKK-----DNPQVNHSQLNESHKEDLQ 89
Db 81 IEKQILKPENRFWHELKQNNPFLYNKKEKKHNLQSIEDTNLQHSQIPSSHPQDLK 140

QY 90 RE-----EHSQKSDSTKDVATVLDKNNI 113
Db 141 OKKDISETKNIOKYNKQEQVQITSQNDL 171

RESULT 19
US-11-196-475-74
/ Sequence 74, Application US/11/196475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Constructs of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ CURRENT FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ PRIOR FILING DATE: 2001-08-07
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 74
/ LENGTH: 700
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match          12.1%; Score 78.5; DB 7; Length 700;
Best Local Similarity 24.3%; Pred. No. 26;
Matches 27; Conservative 29; Mismatches 32; Indels 23; Gaps 5;

QY 36 INLSKDTFTFKPVFKIEEKEENKPT-FDVSKKKDNFPQV-----HSQLNES----H 83
Db 316 LDIQRTVREKLQENINETNKEKNLPKPCDVSSPKVDXQLQIKESLEDLQEQLKETGDN 375

QY 84 RKEDLQREHSQKSD-----STKDVATVLDKNNISSKSTNNPNK 124
Db 376 QKREIEKQTEIKSDEKLLKSKDKASKDGKALDLDR-ELNSKASSKEKSK 425
```

## RESULT 20

```
US-10-793-626-1780
/ Sequence 1780, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1780
/ LENGTH: 1155
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
/ OTHER INFORMATION: amino acid sequence
US-10-793-626-1780
```

```
Query Match          12.1%; Score 78.5; DB 6; Length 1155;
Best Local Similarity 27.3%; Pred. No. 46;
Matches 24; Conservative 12; Mismatches 27; Indels 25; Gaps 2;

QY 53 EKKEEENKPTFDVSKKKDN-----PQVNHSQLNESHKEDLQREHSQK 96
Db 100 EVKKVEAPTTSVSKPANEAVVTNESTKPKTTEAPTVEBSIAETPKTSTTQDDSTEKN 159

QY 97 SDSTKDVATVLDKNNISSKSTNNPNK 124
Db 160 NPSLKD-----NLNSSSTTSKESK 178
```

## RESULT 21

```
US-10-485-517-239
/ Sequence 239, Application US/10485517
/ Publication No. US20050256299A1
/ GENERAL INFORMATION:
/ APPLICANT: University of Sheffield
/ APPLICANT: Biosynexus Incorporated
/ APPLICANT: Foster, Simon
/ APPLICANT: Mond, James
/ TITLE OF INVENTION: Antigenic Polypeptides
/ FILE REFERENCE: P100629WO
/ CURRENT APPLICATION NUMBER: US/10/485,517
/ CURRENT FILING DATE: 2004-02-02
/ PRIOR APPLICATION NUMBER: GB 0118825.9
/ PRIOR FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: GB 0200349.9
/ PRIOR FILING DATE: 2002-01-09
/ NUMBER OF SEQ ID NOS: 424
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 239
/ LENGTH: 405
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-485-517-239
```

```
Query Match          11.9%; Score 77.5; DB 6; Length 405;
Best Local Similarity 25.7%; Pred. No. 17;
Matches 22; Conservative 16; Mismatches 33; Indels 3; Gaps 2;

QY 49 KKIEEKEENKPTFDVSKKKDNFPQVNHSQLN-ESHKEDLQREHSQKSDSTKDVATV 107
Db 51 KAAESTKNELNEATISADNQSSDKVDVQQLNQEDNTKNDQKEMVSSQGNETTSGNKL 110

QY 108 LDKNNISSKSTNN 121
Db 111 IEKESV--QSTTGN 122
```

```
RESULT 22
US-10-873-528-191
; Sequence 191, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-191

Query Match      11.9%; Score 77.5; DB 6; Length 627;
Best Local Similarity 23.4%; Pred. No. 28;
Matches 30; Conservative 26; Mismatches 59; Indels 13; Gaps 4;

Qy 10 CELEKGY--QFDGWEISGEFGKAGYVINLSKDTFIPKVPKIEEKKKEENKPT----- 62
Db 497 GAKQGWHKANDWFFYKTDGSRAGWIKDKQKWFLEKNGQLLVNGKTPGTYVDSSGA 556

Qy 63 --FVSKKKON--POVNHQSQLNESH--KEDLQREHSQKSDSKDVTATVLDKNNISSK 116
Db 557 WLVDVSIKESATIKTTSHEIKESKEVVKOLENKETSQHSVTFNSTSQDLSSTSQSS 616

Qy 117 STNNPNK 124
Db 617 ETSVKNSE 624

RESULT 23
US-11-196-475-70
; Sequence 70, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.9%; Score 77.5; DB 6; Length 627;
Best Local Similarity 23.4%; Pred. No. 28;
Matches 30; Conservative 26; Mismatches 59; Indels 13; Gaps 4;

Qy 10 CELEKGY--QFDGWEISGEFGKAGYVINLSKDTFIPKVPKIEEKKKEENKPT----- 62
Db 497 GAKQGWHKANDWFFYKTDGSRAGWIKDKQKWFLEKNGQLLVNGKTPGTYVDSSGA 556

Qy 63 --FVSKKKON--POVNHQSQLNESH--KEDLQREHSQKSDSKDVTATVLDKNNISSK 116
Db 557 WLVDVSIKESATIKTTSHEIKESKEVVKOLENKETSQHSVTFNSTSQDLSSTSQSS 616

Qy 117 STNNPNK 124
Db 617 ETSVKNSE 624

RESULT 24
US-11-196-475-78
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-78

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ--LN--ESHRKEDLQREHSQKSDSKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGEBEDKELDSKN 432

Qy 103 V-TATVLDKNNISSKSTNNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 25
US-11-188-298-17915
; Sequence 17915, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
```

```
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ--LN--ESHRKEDLQREHSQKSDSKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGEBEDKELDSKN 432

Qy 103 V-TATVLDKNNISSKSTNNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455
```

```
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ--LN--ESHRKEDLQREHSQKSDSKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGEBEDKELDSKN 432

Qy 103 V-TATVLDKNNISSKSTNNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 25
US-11-188-298-17915
; Sequence 17915, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
```



; CURRENT FILING DATE: 2005-09-21  
; PRIOR APPLICATION NUMBER: 60/612,310  
; PRIOR FILING DATE: 2004-09-21  
; NUMBER OF SEQ ID NOS: 88  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 33  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-232-440-33

Query Match 11.7%; Score 76; DB 7; Length 482;  
Best Local Similarity 21.4%; Pred. No. 28;  
Matches 27; Conservative 24; Mismatches 39; Indels 36; Gaps 5;  
QY 24 SGFEGKADGYVINLSKDTFIKVPFKKIEKK-----EENKPTFDVSK 67  
DB 40 SGLSQKEE-----EEDTFIEE--QLSEKLLERQRLHEEWLLREKQAEFRICK 90  
QY 68 KDNQOVNH-----SOLNESHKEDLQREHSQKSDSTKYDTATVLD--KNNSK 116  
DB 91 EKEEAARKQEQERKLQEQWEEQOKREBEERQKQKKEBEALQKMLDQAELENG 150  
QY 117 STNNP 122  
DB 151 TTQMP 156

RESULT 30  
US-10-793-626-2700  
; Sequence 2700, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2700  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2700

Query Match 11.6%; Score 75.5; DB 6; Length 278;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 28; Conservative 16; Mismatches 45; Indels 29; Gaps 4;  
QY 36 INLSKDTFIKVPFKKIEKKEE-----ENKPTFDVSKKKNQVNHSQLNESHKED-- 87  
DB 98 INKCEETNNNDGIESSEDRSTSTTVNDEATFLQSPQDNTLTLTEEVKEPSPVSSN 157  
QY 88 -----LQREHSQKSDTKD-----VTATVLDKNNSK--STNNPNK 124  
DB 158 S8IDTAQPSHTTINREESVQTSNVDVSDPANSKIKESNTSGKEENTIEQPNK 215

RESULT 31  
US-10-793-626-398  
; Sequence 398, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 398  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (404)  
; OTHER INFORMATION: variable amino acid  
US-10-793-626-398

Query Match 11.6%; Score 75.5; DB 6; Length 404;  
Best Local Similarity 29.0%; Pred. No. 25;  
Matches 27; Conservative 18; Mismatches 33; Indels 15; Gaps 5;  
QY 39 SKDTFIKVPFKKIEKKKEENKPTFDVSKKKNQVNH-----HSQLNESHKEDLQRE 92  
DB 161 TKTSTSPFTYLNHSFKSSEVPSAIFGTKRRRIENGVIPEKELND---KEIVQODE 217  
QY 93 --HSQKS-DSTKDVAT---VLDKNNSKSTT 119  
DB 218 VSHSTKSIDASKNVSNNDNIEKNQKQKQTT 250

RESULT 32  
US-10-510-903-22  
; Sequence 22, Application US/10510903  
; Publication No. US20060051755A1  
; GENERAL INFORMATION:  
; APPLICANT: Hitoshi, Yasumichi  
; APPLICANT: Jenkins, Yonchu  
; APPLICANT: Rigel Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods of Assaying for Cell Cycle Modulators  
; FILE REFERENCE: 021044-003310PC  
; CURRENT APPLICATION NUMBER: US/10/510,903  
; CURRENT FILING DATE: 2004-10-08  
; PRIOR APPLICATION NUMBER: US 10/123,568  
; PRIOR FILING DATE: 2002-04-15  
; PRIOR APPLICATION NUMBER: US 10/123,731  
; PRIOR FILING DATE: 2002-04-15  
; PRIOR APPLICATION NUMBER: US 60/373,366  
; PRIOR FILING DATE: 2002-04-16  
; PRIOR APPLICATION NUMBER: WO PCT/US03/11867  
; PRIOR FILING DATE: 2003-04-15  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 558  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21  
; OTHER INFORMATION: (DDX21) (RNA helicase II)  
US-10-510-903-22

Query Match 11.6%; Score 75.5; DB 6; Length 558;  
Best Local Similarity 26.9%; Pred. No. 36;  
Matches 32; Conservative 18; Mismatches 40; Indels 29; Gaps 6;  
QY 29 KQDAGYVNLKDTFIK--PVFKIEKKKEENKPTFDVS-----KXK 69  
DB 6 RSDAG----LESDTAMKGETLRQTEKKEKPKSKDTEIEAEETVFPKAKQVKK 61  
QY 70 DNPQVNHSQLNESHKEDLQREHSQK--SUSTKDV--TATVLDKNNSKSTTNNPNK 124  
DB 62 AEP--SEYDNNSPKSKAKKKEPSSQNDISPTKSLRKKKPKIEKKVSVSKTKVKNE 118

## RESULT 33

US-11-185-924-16  
; Sequence 16, Application US/11185924  
; Publication No. US20060078945A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher et al., Larry  
; TITLE OF INVENTION: Complex Formed by Small Integrin-Binding Ligand,  
; TITLE OF INVENTION: N-Linked Glycoproteins (SIBLINGS) and Factor H  
; FILE REFERENCE: 4239-61301-02  
; CURRENT APPLICATION NUMBER: US/11/185,924  
; CURRENT FILING DATE: 2005-07-19  
; PRIOR APPLICATION NUMBER: 09/958,617  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: PCT/US00/09349  
; PRIOR FILING DATE: 2000-04-09  
; PRIOR APPLICATION NUMBER: 60/128,468  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-185-924-16

Query Match 11.5%; Score 75; DB 7; Length 513;  
Best Local Similarity 23.5%; Pred. No. 37;  
Matches 23; Conservative 27; Mismatches 34; Indels 14; Gaps 4;

QY 29 KKDAGVNLKDTIKPVFKIEKKEEENKPTFDVSKKK-DNPQVNHSQLNE-----S 82  
DB 385 ESDSSHTLHSHKS-----ESREQADSESESLENFSESPEDENSSSQEGLOSHS 437  
QY 83 HRKDLQREHSQKSDS-TKQVTATVLDKQNNISSKSTT 119  
DB 438 SSAESQSESHSEEDSDSQDSRSKSDNSTESKSS 475

## RESULT 34

US-11-096-568A-32043  
; Sequence 32043, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32043  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(313)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592023  
US-11-096-568A-32043

Query Match 11.4%; Score 74.5; DB 7; Length 313;  
Best Local Similarity 22.0%; Pred. No. 23;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;

QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
DB 196 ELYVFPKLYGVGVELWQVKSGLFDNLVLS-----DDPEYAKKLAETWGHKDKAEKAAF 250  
QY 49 KKIEKKEEENKPTFDVSKKKNPQVNHSQLNESHRKDLQREHSQKSDSTKQVTAT 106  
DB 251 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEETKEAETKEAET 306

## RESULT 35

US-11-096-568A-32042  
; Sequence 32042, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32042  
; LENGTH: 425  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(425)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592022  
US-11-096-568A-32042

Query Match 11.4%; Score 74.5; DB 7; Length 425;  
Best Local Similarity 22.0%; Pred. No. 32;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;

QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
DB 308 ELYVFPKLYGVGVELWQVKSGLFDNLVLS-----DDPEYAKKLAETWGHKDKAEKAAF 362  
QY 49 KKIEKKEEENKPTFDVSKKKNPQVNHSQLNESHRKDLQREHSQKSDSTKQVTAT 106  
DB 363 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEETKEAETKEAET 418

## RESULT 36

US-11-096-568A-32041  
; Sequence 32041, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32041  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(434)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592021  
US-11-096-568A-32041

Query Match 11.4%; Score 74.5; DB 7; Length 434;  
Best Local Similarity 22.0%; Pred. No. 33;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;

QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
DB 317 ELYVFPKLYGVGVELWQVKSGLFDNLVLS-----DDPEYAKKLAETWGHKDKAEKAAF 371  
QY 49 KKIEKKEEENKPTFDVSKKKNPQVNHSQLNESHRKDLQREHSQKSDSTKQVTAT 106  
DB 372 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEETKEAETKEAET 427

## RESULT 37

Search completed: April 24, 2006, 15:44:46  
Job time : 10.7092 secs



**THIS PAGE BLANK (user)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 62.1387 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773  
Perfect score: 651  
Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	4	US-10-067-385-8
2	651	100.0	2119	3	US-09-769-744A-28
3	651	100.0	2140	4	US-10-282-122A-73670
4	651	100.0	2140	5	US-10-472-928-1180
5	648	99.5	637	5	US-10-617-320-3169
6	615	94.5	117	3	US-09-765-272-68
7	615	94.5	117	6	US-11-106-649-68
8	106	16.3	188	5	US-10-691-672A-7
9	101.5	15.6	564	6	US-11-097-143-12723
10	101	15.5	169	5	US-10-691-672A-2
11	101	15.5	647	5	US-10-691-672A-3
12	98	15.1	665	3	US-09-820-843A-107
13	96.5	14.8	707	4	US-10-282-122A-52942
14	93	14.3	869	4	US-10-437-963-12282
15	93	14.3	1529	5	US-10-732-923-8762
16	92.5	14.2	2468	4	US-10-755-889-615
17	92.5	14.2	2468	5	US-10-489-740-216
18	92.5	14.2	2519	5	US-10-450-763-46995
19	92	14.1	1005	4	US-10-437-963-187665
20	91	14.0	470	5	US-10-739-930-6262
21	89.5	13.7	511	4	US-10-289-762-509
22	88	13.5	225	4	US-10-032-585-7829
23	88	13.5	815	5	US-10-496-905-23
24	88	13.5	1980	5	US-10-482-834A-144
25	87.5	13.4	903	4	US-10-282-122A-52328
26	87.5	13.4	1702	3	US-09-839-996-5
27	87.5	13.4	1702	4	US-10-080-505-5

28	87.5	13.4	1702	4	US-10-645-655-5	Sequence 5, Appli
29	87.5	13.4	1702	4	US-10-687-046-5	Sequence 5, Appli
30	87	13.4	593	4	US-10-424-599-174814	Sequence 174814, A
31	86.5	13.3	1373	5	US-10-732-923-16976	Sequence 16976, A
32	86	13.2	145	4	US-10-424-599-149572	Sequence 149572, A
33	86	13.2	835	4	US-10-425-115-238086	Sequence 238086, A
34	86	13.2	1075	4	US-10-437-963-187664	Sequence 187664, A
35	86	13.2	1726	5	US-10-942-522-6	Sequence 6, Appli
36	86	13.2	1980	4	US-10-474-778-6	Sequence 6, Appli
37	86	13.2	1980	5	US-10-482-834A-150	Sequence 150, App
38	86	13.2	1980	5	US-10-942-522-8	Sequence 8, Appli
39	85.5	13.1	884	4	US-10-032-585-7212	Sequence 7212, Ap
40	85.5	13.1	982	4	US-10-425-115-293706	Sequence 293706, A
41	85	13.1	316	4	US-10-424-599-180899	Sequence 180899, A
42	85	13.1	465	4	US-10-724-972A-5599	Sequence 5599, Ap
43	85	13.1	728	4	US-10-282-122A-70944	Sequence 70944, A
44	85	13.1	728	4	US-10-724-972A-4389	Sequence 4389, Ap
45	85	13.1	778	4	US-10-724-972A-5663	Sequence 5663, Ap

ALIGNMENTS

RESULT 1  
US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 651;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.6e-51;		
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVNLKDTFIKPVFKLEEKEENK	60	
Db	650	EDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVNLKDTFIKPVFKLEEKEENK	709	
Qy	61	PTFDVSKKKQPNVHNSQLNESHKRDILQREHSQKSDSTKDVATVLDKNNISSKSTTN	120	
Db	710	PTFDVSKKKQPNVHNSQLNESHKRDILQREHSQKSDSTKDVATVLDKNNISSKSTTN	769	
Qy	121	NNPK 124		
Db	770	NNPK 773		

RESULT 2  
US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 651; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.5e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 60
    |||||
Db 1962 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 2021

QY 61 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120
    |||||
Db 2022 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 2081

QY 121 NPNK 124
    |||||
Db 2082 NPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 651; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 60
    |||||
Db 1983 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 2042

QY 61 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120
    |||||
Db 2043 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 2102

QY 121 NPNK 124
    |||||
Db 2103 NPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/104729283
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 651; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 60
    |||||
Db 1983 EDFILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINLSKDTFIKPVFKKIEEKKKEENK 2042

QY 61 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120
    |||||
Db 2043 PTFDVSKKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 2102

QY 121 NPNK 124
    |||||
Db 2103 NPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORGANISM: Streptococcus pneumoniae
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169

Query Match          99.5%; Score 648; DB 5; Length 637;
Best Local Similarity 99.2%; Pred. No. 1.1e-50;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVPKKEEKKEENK 60
Db 480 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVPKKEEKKEENK 539
Qy 61 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTTATVLDKNNISKSTTN 120
Db 540 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTTATVLDKNNISKSTTN 599
Qy 121 NPNK 124
Db 600 NPNK 603

RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US2002006154A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match          94.5%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVPKKEEKKEENKPTFDVSK 67
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVPKKEEKKEENKPTFDVSK 60
Qy 68 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTTATVLDKNNISKSTTNPNK 124
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTTATVLDKNNISKSTTNPNK 117

RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

2

Qy 120 NN 121  
Db 125 NN 126

RESULT 11  
US-10-691-672A-3  
; Sequence 3, Application US/10691672A  
; Publication No. US20050112133A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
; FILE REFERENCE: 02356.0085  
; CURRENT APPLICATION NUMBER: US/10/691,672A  
; CURRENT FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 3  
; LENGTH: 647  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; NAME/KEY: SITE  
; LOCATION: (1)..(647)  
; OTHER INFORMATION: GLURP MSP3 fusion protein  
US-10-691-672A-3

Query Match 15.5%; Score 101; DB 5; Length 647;  
Best Local Similarity 25.4%; Pred. No. 1.2;  
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

Qy 20 GWEISGP--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKQNP 72  
Db 489 GWEFGGGVPEHKKEENMLSHLYVSSKKENISKENDDDVLDE-KGEAEATETEELEEENK 547  
Qy 73 QVNHSQLN-----ESHKEDLQREHSQKSDSTKVDTATVLDKNNISSKSTT 119  
Db 548 EETSEISEDEEEEEKEEENKKEQKEQENNDQKKDMEA-----QNLISKNQN 602

Qy 120 NN 121  
Db 603 NN 604

RESULT 12  
US-09-820-843A-107  
; Sequence 107, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: 063915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 107  
; LENGTH: 665  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: g1|3845248  
US-09-820-843A-107

Query Match 15.1%; Score 98; DB 3; Length 665;  
Best Local Similarity 27.1%; Pred. No. 2.3;  
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

Qy 13 EKGQFDGWEI--SGPEKKDAGYVINLSKDTPIKPVFKKIEKKKE-----EENK 60  
Db 158 EKGQKQ----DISNSAENKKD-----VKGVKLEEKKEKKKEKISDDHKVEENK 201  
Qy 61 PTFD----VSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKVDTATVLDKNNISS 115  
Db 202 KSDDHKEENKSDDHKEENKSDDHKEIEVKVKEHEEDEEE-----DKCEKKS 252

Qy 116 KSTNNPNK 124  
Db 253 ENKKNENK 261

RESULT 13  
US-10-282-122A-52942  
; Sequence 52942, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zvakind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 52942  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (6)..(6)  
; OTHER INFORMATION: X=any amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (18)..(18)  
; OTHER INFORMATION: X=any amino acid

```
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (359)..(359)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (385)..(385)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (388)..(388)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (396)..(396)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (404)..(404)
; OTHER INFORMATION: X=any amino acid
;
US-10-282-122A-52942
```

```
Query Match 14.8%; Score 96.5; DB 4; Length 707;
Best Local Similarity 25.2%; Pred. No. 3.4;
Matches 34; Conservative 25; Mismatches 53; Indels 23; Gaps 4;

Qy 9 KBELEKGYQFGWEISGPEKKDAGVINLSKDTPIKPVFKKIEK----- 54
Db 495 KEEIDSNQDIGVV---EDKDTTDEYDSNKEIDIEPENKSKKKAKLFGPIKKDNEV 551
Qy 55 -KEEN-----KTPFVSKKONPQVNHSQLSHRKEDLQREHSQKSDSKDVTATVLD 109
Db 552 EQEENLNDISPDITLDPVENNQVSKSEIEQNELKE-IKOBEPQHIERSVKIEKPI 610
Qy 110 KNNISSKSTTNNPNK 124
Db 611 NNNLDEKVSSNNESK 625
```

```
RESULT 14
US-10-437-963-122282
```

```
; Sequence 122282, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRM4530_25224C.1.1.pep
US-10-437-963-122282

Query Match 14.3%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 9.2;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

Qy 64 DVSKKONPQVNHSQLSHRKEDLQREHSQKSDSKDVTATVLDKNN 112
Db 617 DASKRKDHQSEGNNL--SHRDEDPTRKKKQKTNATSDACAQEVWTEKNN 665
```

```
RESULT 15
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762
```

```
Query Match 14.3%; Score 93; DB 5; Length 1529;
Best Local Similarity 27.2%; Pred. No. 18;
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

Qy 37 NLSKDT-----FIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQL-----NES 82
Db 356 NIISDTQCIKPIKVINSEYKNEKKNEKK-----NEKINTIHYSESISKNSDNEQ 409
Qy 83 H-----RKEDLQREHSQKSDSKDVTATVLDKNN--ISSKSTT 119
Db 410 HPFLSKLRNVKKEKKKKKKTKIKTVIAQKNKTVIAQKNKT 452
```

```
RESULT 16
US-10-755-889-615
; Sequence 615, Application US/10755889
```



```
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 615
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-615

Query Match      14.2%; Score 92.5; DB 4; Length 2468;
Best Local Similarity 31.0%; Pred. No. 36;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKDTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92
Db 638 VKKETKVKP-----EDKKEEKKPKKEVAKKEDKTPI---KKEEKKPKKEEVKKEVKKKEIK 689

QY 93 HSQKSDSTKDV 103
Db 690 KEEKKEPKKEV 700

RESULT 17
US-10-489-740-216
; Sequence 216, Application US/10489740
; Publication No. US20050112574A1
; GENERAL INFORMATION:
; APPLICANT: Biomedics Limited
; TITLE OF INVENTION: P9
; FILE REFERENCE: Angiogenesis PCT
; CURRENT APPLICATION NUMBER: US/10/489,740
; CURRENT FILING DATE: 2004-03-15
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 216
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-489-740-216

Query Match      14.2%; Score 92.5; DB 5; Length 2468;
Best Local Similarity 31.0%; Pred. No. 36;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKDTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92
Db 638 VKKETKVKP-----EDKKEEKKPKKEVAKKEDKTPI---KKEEKKPKKEEVKKEVKKKEIK 689

QY 93 HSQKSDSTKDV 103
Db 690 KEEKKEPKKEV 700

RESULT 18
US-10-450-763-46995
; Sequence 46995, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
```

```
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 46995
; LENGTH: 2519
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1040)..(1091)
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1919)..(2122)
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by PFam,
; OTHER INFORMATION: accession name MAP1B_neuraxin, E-value=1.9e-59, PFam score of 19.54
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2519)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-46995

Query Match      14.2%; Score 92.5; DB 5; Length 2519;
Best Local Similarity 31.0%; Pred. No. 37;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKDTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92
Db 689 VKKETKVKP-----EDKKEEKKPKKEVAKKEDKTPI---KKEEKKPKKEEVKKEVKKKEIK 740

QY 93 HSQKSDSTKDV 103
Db 741 KEEKKEPKKEV 751

RESULT 19
US-10-437-963-187665
; Sequence 187665, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187665
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84346C.1.pep
US-10-437-963-187665

Query Match      14.1%; Score 92; DB 4; Length 1005;
Best Local Similarity 26.4%; Pred. No. 14;
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;
```

Qy 25 GPEGKADAGYVNLKDTFKIPVFKKIBEKKEEENKPTFDVSKK-KONPQV--NHSQLE 81  
Db 35 GSEKEMSGKNKSIKETGTGQSKELQKSKSRKSTKDKSKKDKMTQVPTNAEBPHK 94  
Qy 82 SH-----RK-----DLQREHSQKSDTKDVTATVLDKNNISSKST 118  
Db 95 EYTTVKIRKESRTDSSIEQVIGTSSIQEMETNEQVSKDTSKDMTQVPAEAGIRKEYT 154  
Qy 119 T 119  
Db 155 T 155  
RESULT 20  
US-10-739-930-6262  
; Sequence 6262, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377) B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; NUMBER OF SEQ ID NOS: 12-18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6262  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270\_1.p  
US-10-739-930-6262  
Query Match 14.0%; Score 91; DB 5; Length 470;  
Best Local Similarity 19.3%; Pred. No. 6.7;  
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;  
Qy 8 YKGELEKGVQPDGWEISGPE-----GKKDAGYVNLKDTFKIPVFKKIBEKKEEN 59  
Db 100 YVQDLARRIYDE-EATQSQAQRIDHPNQKQVGVGITEKAFENSPLEETSHRVNDKRN 158  
Qy 60 KPTFDVSKKON-----PQVNHSQLNE-----SHRKEDLQREHS 94  
Db 159 QKNFPAKSSSENAVSRVSGADHKRAEYVMGKPMENRDQVRQTESAEKSHRKENVTSEKP 218  
Qy 95 QKSDTKDVTATVLDKNNISSKSTTNPNK 124  
Db 219 RDQGVKKTEAKDKRNKEKKEETESINK 248  
RESULT 21  
US-10-289-762-509  
; Sequence 509, Application US/10289762  
; Publication No. US20040006218A1  
; GENERAL INFORMATION:  
; APPLICANT: Grifffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; FILE REFERENCE: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/10/289,762  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 509  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-10-289-762-509  
Query Match 13.7%; Score 89.5; DB 4; Length 511;  
Best Local Similarity 24.5%; Pred. No. 10;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 44 IKPVFKKIBEKKEENKPTFD-----VSKKDNQVNHSQLNESHK 85  
Db 95 VKGVFKTTPQAREVSSPRLPVHQGLGEGFDRIQKSENFEADLGKKRSYSD 154  
Qy 86 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 119  
Db 155 GDLDRVGHDSNEDSTEDSR-----EGCEPSSKSSS 185  
RESULT 22  
US-10-032-585-7829  
; Sequence 7829, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-959  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7829  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7829  
Query Match 13.5%; Score 88; DB 4; Length 225;  
Best Local Similarity 28.0%; Pred. No. 5.2;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
Qy 16 YQPDGWEISGFEKGDAGYVNLKDTFKIPVFKK-----IEKKKEENKPTFDVSKK-- 69  
Db 80 YDDDDDFEGFESSNGAAKELNLSQAIKEWKQRDLIEEREKLNKKKEEIEERAKS 139  
Qy 70 --DNPQVNHSQLNESHKEDLQREH--SQKSDTKDVTATVLDKNN 112  
Db 140 TIDDFYENYKSRDNHOKBILSEQEKFTSKRDDFLK--RGLWDRVN 184  
RESULT 23  
US-10-496-905-23  
; Sequence 23, Application US/10496905  
; Publication No. US20050192215A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Malabika  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Zhao, Qing  
; APPLICANT: Xu, Chongjun  
; APPLICANT: Mulero, Julio J.  
; APPLICANT: Boyle, Bryan J.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYNUC  
; FILE REFERENCE: HYS-BICIP/US  
; CURRENT APPLICATION NUMBER: US/10/496,905  
; CURRENT FILING DATE: 2004-05-26  
; PRIOR APPLICATION NUMBER: US 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: PCT/US00/35017  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: PCT/US00/34263  
; PRIOR FILING DATE: 2000-12-26  
; PRIOR APPLICATION NUMBER: US 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; PRIOR APPLICATION NUMBER: US 09/560,875



QY	36	INLSKDTFTKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHS-	94
Db	1296	INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI	1355
QY	95	--QKSDSTKDVATVLDKNNISSKSTNNPNK 124	
Db	1356	SQPQTSABETTAASTDETTIADNSKRSKPNR 1387	
RESULT 28			
US-10-645-655-5			
Sequence 5, Application US/10645655			
Publication No. US20040063908A1			
GENERAL INFORMATION:			
APPLICANT: St. Geme III, Joseph W.			
Falkow, Stanley			
TITLE OF INVENTION: Haemophilus Adherence and Penetration			
NUMBER OF SEQUENCES: 9			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert			
STREET: 4 Embarcadero Center, Suite 3400			
CITY: San Francisco			
STATE: California			
COUNTRY: United States			
ZIP: 94111-4187			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patent In Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/10/645,655			
FILING DATE: 20-Aug-2003			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US/08/296,791			
FILING DATE: 25-AUG-1994			
ATTORNEY/AGENT INFORMATION:			
NAME: Trecartin, Richard F.			
REGISTRATION NUMBER: 31,801			
REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS			
TELEPHONE: (415) 781-1989			
TELEFAX: (415) 398-3249			
TELEX: 910 277299			
INFORMATION FOR SEQ ID NO: 5:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 1702 amino acids			
TYPE: amino acid			
TOPOLOGY: unknown			
SEQUENCE DESCRIPTION: SEQ ID NO: 5:			
US-10-645-655-5			
Query Match	13.4%	Score 87.5; DB 4; Length 1702;	
Best Local Similarity	27.2%	Pred. No. 65;	
Matches	25; Conservative	14; Mismatches	50; Indels 3; Gaps 1;
QY	36	INLSKDTFTKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHS-	94
Db	1296	INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI	1355
QY	95	--QKSDSTKDVATVLDKNNISSKSTNNPNK 124	
Db	1356	SQPQTSABETTAASTDETTIADNSKRSKPNR 1387	
RESULT 27			
US-10-080-505-5			
Sequence 5, Application US/10080505			
Publication No. US20030073166A1			
GENERAL INFORMATION:			
APPLICANT: St. Geme, Joseph W.			
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS			
FILE REFERENCE: A-59941-1/RPT/DCP/DHR			
CURRENT APPLICATION NUMBER: US/10/080,505			
CURRENT FILING DATE: 2002-02-22			
PRIOR APPLICATION NUMBER: US 08/296,791			
PRIOR FILING DATE: 1994-10-25			
PRIOR APPLICATION NUMBER: US 09/839,996			
PRIOR FILING DATE: 2001-04-20			
NUMBER OF SEQ ID NOS: 58			
SOFTWARE: Patent In version 3.1			
SEQ ID NO 5			
LENGTH: 1702			
TYPE: PRT			
ORGANISM: Haemophilus influenzae			
US-10-080-505-5			
Query Match	13.4%	Score 87.5; DB 4; Length 1702;	
Best Local Similarity	27.2%	Pred. No. 65;	
Matches	25; Conservative	14; Mismatches	50; Indels 3; Gaps 1;





Qy	17	QPDGWEISGEGKGDAGVIVNLSXDTIPKVPFKKBEKKEBENKPTFDVSKGONPQVNH	76
Db	993	EMNNQLSVIRIKKGAVT--KLKVHAFMOAHFK--QREADEVXPLDELVEKGCANTAH	1048
Qy	77	SQLMESHREKEDLQREHSQKSDSTKVATATVLDKNNISSSKSTNNPN	123
Db	1049	TGA--DIHRNGDFQNGNGTTSIGISSVEKVIIDEDHM---SFINNPN	1091

```

RESULT 37
US-10-482-834A-150
; Sequence 150, Application US/10482834A
; Publication No. US2005007464A1
; GENERAL INFORMATION:
; APPLICANT: Mulley, John Charles
; APPLICANT: Harkin, Louise Anne
; APPLICANT: Dibbens, Michelle
; APPLICANT: Wallace, Robyn
; APPLICANT: Phillips, Hillary Anny
; APPLICANT: Heron, Sara Elizabeth
; APPLICANT: Berkovic, Samuel Frank
; APPLICANT: Scheffer, Ingrid Eileen
; APPLICANT: Bionomics Limited
; TITLE OF INVENTION: MUTATIONS IN ION CHANNELS
; FILE REFERENCE: 1386/17
; CURRENT APPLICATION NUMBER: US/10/482,834A
; CURRENT FILING DATE: 2004-01-02
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 1980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-482-834A-150

```

	Query Match	Best Local Similarity	13.28; Score 86; DB 5; Length 1980;
	Matches	29; Conservative	21; Mismatches 49; Indels 8; Gaps 4;
Qy	17	QFDGWEISGEGKKDAGVYNLSKDTTIPKVPFKIIEKQGEENKPTFDVSKKGNPQVNH	76
Db	993	EMNNLQVSVIRIKKGVAWT-KLKVHAPMAHFQCE--ADEVKPLDELKVCANCIAH	1048
Qy	77	SQLNESHKEDLQREHSQKSDSTKQVATATVLDKNNISKSSTNNPN	123
Db	1049	TGA-DIHRNGDFQKNGTGTSGIGSSVEKVIIDEDHM---SFINNPN	1091

RESULT 38  
US-10-942-522-8  
; Sequence 8, Application US/10942522  
; Publication No. US20050112633A1  
; GENERAL INFORMATION:  
; APPLICANT: Armour, Christopher D.  
; APPLICANT: Castle, John C.  
; APPLICANT: Garrett-Englele, Philip W.  
; APPLICANT: Kan, Zhengyan  
; APPLICANT: Loerch, Patrick M.  
; APPLICANT: Teinoremnas, Nicholas P.  
; TITLE OF INVENTION: ALTERNATIVELY SPLICED ISOFORMS OF SODIUM  
; TITLE OF INVENTION: CHANNEL, VOLTAGE GATED, TYPE VIII, ALPHA (SCN8A) (SCN8A)  
; FILE REFERENCE: RS0214  
; CURRENT APPLICATION NUMBER: US/10/942,522  
; CURRENT FILING DATE: 2004-09-16  
; PRIOR APPLICATION NUMBER: US 60/503,694  
; PRIOR FILING DATE: 2003-09-17  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1980  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-942-522-8

	Query Match	13.2k;	Score 86;	DB 5;	Length 1980;
	Best Local Similarity	26.2k;	Pred. No. 1.1e+02;		
	Matches 28;	Conservative	Mismatches 48;	Indels	Gaps 4;
Qy	17	QFGWEISGPEGKDGAGVNLKDTTFIKVPFKKIEBKKEENKPTFDVSKGKDNPPVNH	76		
Db	993	EMNLQGVSVIRIKKGVAWT-KLKVHAFMQAHFK---QREADVKPLDELYEKKANCIAANH	1048		
Qy	77	SQLNESHKEDLQREEHSSQKSDSTKDVATVLVDKNNISSTSTNNPN	123		
Db	1049	TGA-DIHRNGDFPKNGNGTTSIGSSVEKVIIDEDHM---SFNNPN	1091		

```

RESULT 39
US-10-032-585-7212
; Sequence 7212, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7212
; LENGTH: 884
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7212

```

	Query Match	13.1%	Score 85.5;	DB 4;	Length 884;
	Best Local Similarity	25.0%;	Pred. No. 46;		
	Matches	26;	Conservative 23;	Mismatches 46;	Indels 9; Gaps 3;
Qy	24	SGFEGKDAGYVNLGNKDTFIKPVPKKIE---	EKGEEENKPTPDVSK--	KKNPQNHSQ	78
Db	14	SGGEVNNTNQTSELSKEELKRRLRQLGLWLRKQ	QQQQQEVEVQNKA	KKTEDSTNNTS	73
Qy	79	LNESHRKEDIQRREE---	HSQKSDSTKDVTATVL	DKNNISSKST	118
Db	74	TSEVKKSROORIEEWKRALOKOACAKETTTTTIKK	VHSHSOFTT		117

```

RESULT 40
US-10-425-115-293706
; Sequence 293706, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Thomas J.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 293706
; LENGTH: 982
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(982)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577 30943C.1.bep

```



US-10-425-115-293706

```
Query Match      13.1%; Score 85.5; DB 4; Length 982;
Best Local Similarity 26.7%; Pred. No. 52;
Matches 35; Conservative 18; Mismatches 49; Indels 29; Gaps 5;

QY      20 GWEISGFEGKKDAGYVINLSKDTFIKPVFKK-----IEKKKEENKPTFDVSKKKON-PQV 74
Db      110 GSRLSKFEKKCDGFLLEYSSNSKAYRVFNKTHGIIIEAYDVEFPDKTNGSQDESDNLDDV 169

QY      75 NHSQL-----NESHKED-----LOREEH-SQKSDSTKDVTTATVLDK 110
Db      170 GGTQLINAMKTAIGEIKPKEDDENSVVVIPSSTLNEDHQSOQHDETMTHDQGTSR 229

QY      111 NNISKSSTNN 121
Db      230 HSVPPNASTSN 240
```

Search completed: April 24, 2006, 15:43:31  
Job time : 64.1387 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 19.2796 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773  
Perfect score: 651  
Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCITUS-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	2	US-09-590-991-8
2	648	99.5	637	2	US-09-107-433-3169
3	648	99.5	2138	2	US-09-583-110-5274
4	615	94.5	117	2	US-08-961-083-68
5	615	94.5	117	2	US-09-536-784-68
6	615	94.5	117	2	US-09-765-271-68
7	615	94.5	117	2	US-09-765-272A-68
8	96	14.7	347	2	US-09-248-796A-16224
9	95	14.6	348	2	US-09-538-092-1316
10	92.5	14.2	2468	2	US-09-976-594-726
11	92.5	14.2	2468	2	US-09-538-092-1135
12	92.5	14.2	2522	2	US-09-949-016-10237
13	89.5	13.7	511	2	US-09-198-452A-509
14	89.5	13.7	511	2	US-09-438-185A-475
15	87.5	13.4	1702	2	US-08-296-791-5
16	87.5	13.4	1702	2	US-09-839-996-5
17	87.5	13.4	1702	2	US-10-080-505-5
18	87.5	13.4	1702	2	US-10-645-655-5
19	87.5	13.4	1702	4	PCT-US95-10661A-5
20	86	13.2	243	2	US-09-248-796A-20306
21	86	13.2	1989	2	US-09-949-016-10076
22	85.5	13.1	109	2	US-09-248-796A-24668
23	85	13.1	465	2	US-09-134-001C-3856
24	85	13.1	472	2	US-09-710-279-658
25	85	13.1	720	2	US-09-710-279-2058
26	85	13.1	728	2	US-09-134-001C-4968
27	85	13.1	746	2	US-09-710-279-652

28	85	13.1	778	2	US-09-134-001C-3868
29	85	13.1	1976	2	US-09-024-020B-9
30	85	13.1	1976	2	US-09-425-043-9
31	85	13.1	1978	2	US-09-024-020B-3
32	85	13.1	1978	2	US-09-425-043-3
33	85	13.1	1988	2	US-09-024-020B-4
34	85	13.1	1988	2	US-09-425-043-4
35	84.5	13.0	654	2	US-10-172-502-10
36	84	12.9	817	2	US-09-248-796A-20276
37	83.5	12.8	461	2	US-09-949-016-8508
38	83.5	12.8	476	2	US-09-248-796A-15008
39	83	12.7	278	2	US-09-949-016-10508
40	82.5	12.7	243	2	US-09-248-796A-21334
41	82	12.6	1024	2	US-09-270-767-44973
42	81	12.4	582	2	US-09-147-405B-13
43	81	12.4	593	2	US-09-147-405B-11
44	81	12.4	1092	2	US-09-147-405B-15
45	80	12.3	232	2	US-09-248-796A-23825

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%	Score 651;	DB 2;	Length 773;
Best Local Similarity	100.0%	Pred. No. 1.3e-62;		
Matches 124;	Conservative	0;	Mismatches	0;
	Indels	0;	Gaps	0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFGKQDAGVYINLSKDTPIKVPVKLEKKKEENK	60	
Db	650	EDFILPVYKGELEKGYQFDGWEISGFGKQDAGVYINLSKDTPIKVPVKLEKKKEENK	709	
Qy	61	PTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDTKDVTATVLDKNNISSKSTTN	120	
Db	710	PTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDTKDVTATVLDKNNISSKSTTN	769	
Qy	121	NPNK 124		
Db	770	NPNK 773		

RESULT 2

US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; FOR DIAG



Db 1 YGELKGYQPDGWEISGFEKGDAGYVNLKDTFKPVPFKIEEKEENKPTFDVSK 60  
Qy 68 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124  
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/536,784  
; FILING DATE: 30-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-536-784-68

Query Match 94.5%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.1e-60;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 YGELKGYQPDGWEISGFEKGDAGYVNLKDTFKPVPFKIEEKEENKPTFDVSK 67  
Db 1 YGELKGYQPDGWEISGFEKGDAGYVNLKDTFKPVPFKIEEKEENKPTFDVSK 60  
Qy 68 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124  
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,271  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 09/536,784  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-271-68

Query Match 94.5%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.1e-60;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 YGELKGYQPDGWEISGFEKGDAGYVNLKDTFKPVPFKIEEKEENKPTFDVSK 67  
Db 1 YGELKGYQPDGWEISGFEKGDAGYVNLKDTFKPVPFKIEEKEENKPTFDVSK 60  
Qy 68 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124  
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 7

US-09-765-272A-68  
; Sequence 68, Application US/09765272A  
; Patent No. 6929930  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and  
; NUMBER OF SEQUENCES: 454  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: Dell Latitude C610  
; OPERATING SYSTEM: Windows 2000  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:

```
/
/ APPLICATION NUMBER: US/09/765,272A
/ FILING DATE: 22-Jan-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/961,083
/ FILING DATE: OCT-30-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lin J. Hymel
/ REGISTRATION NUMBER: 45,414
/ REFERENCE/DOCKET NUMBER: PB340P2C2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 610-5790
/ TELEFAX: (301) 309-8439
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match          94.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 9.1e-60;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YGELEKGYQPDGWEISGEGKDGAGYVINLSKDTFIKPVFKKIEKKEENKPTFDVSK 67
Db 1 YGELEKGYQPDGWEISGEGKDGAGYVINLSKDTFIKPVFKKIEKKEENKPTFDVSK 60

QY 68 KKDNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTNNPK 124
Db 61 KKDNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTNNPK 117

RESULT 8
US-09-248-796A-16224
/ Sequence 16224, Application US/09248796A
/ Patent No. 6747137
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstock et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ CURRENT FILING DATE: 1999-02-12
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ PRIOR FILING DATE: 1998-08-13
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 16224
/ LENGTH: 347
/ TYPE: PRT
/ ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match          14.7%; Score 96; DB 2; Length 347;
Best Local Similarity 23.3%; Pred. No. 0.026;
Matches 28; Conservative 22; Mismatches 46; Indels 24; Gaps 3;

QY 11 ELEKGYQPDGWEISGEGKDGAGYVINLSKDTFIKPVFKKIEKKEENKPTFDVSKKD 70
Db 238 DLESESEIGWE-----DDLLENTRTGPFVKTLDLQRE-----WKAKEKQA 279

QY 71 NPQVNHSQLN-----ESHRKEDLQREHSQKSDSTKDVATVLDKNNISKSTTNNPK 124
Db 280 NPKKEENLNQRPVAKQKQPNSTKKQTKQKQTKKITKPKTSKRMLEGISTNSIINK 339

RESULT 9
US-09-538-092-1316
```

```
/
/ Sequence 1316, Application US/09538092
/ Patent No. 6753314
/ GENERAL INFORMATION:
/ APPLICANT: Giot, Loic
/ APPLICANT: Mansfield, Traci A.
/ TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
/ FILE REFERENCE: 15966-542
/ CURRENT APPLICATION NUMBER: US/09/538,092
/ CURRENT FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: 60/127,352
/ PRIOR FILING DATE: 1999-04-01
/ PRIOR APPLICATION NUMBER: 60/178,965
/ PRIOR FILING DATE: 2000-02-01
/ NUMBER OF SEQ ID NOS: 1387
/ SOFTWARE: CurapatSeqformatter Version 0.9
/ SEQ ID NO 1316
/ LENGTH: 348
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (0)...(0)
/ OTHER INFORMATION: Polypeptide Accession Number Q14093
US-09-538-092-1316

Query Match          14.6%; Score 95; DB 2; Length 348;
Best Local Similarity 30.9%; Pred. No. 0.033;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 9 KGELEKGYQPDGWEISGEGKDGAGYVINLSKDTFIKPVFKKIEKKEENKPTF---DV 65
Db 205 ESEGEKG---GTEKDSKKGKDS-----KKGKDSAIELQAVKADKKDCKDANKGDE 256

QY 66 SK--KKDNPVNHSQLN-----ESHRKEDLQREHSQKSDSTKD---VTATVLDKNNI 113
Db 257 SKDAKDAKEIKGKDKKPKPSSTDSDSKDVKKX---SKDATKDAKVAKDKTEKESA 313

QY 114 SSK 116
Db 314 DSK 316

RESULT 10
US-09-976-594-726
/ Sequence 726, Application US/09976594
/ Patent No. 6673549
/ GENERAL INFORMATION:
/ APPLICANT: Furness, Michael
/ APPLICANT: Buchbinder, Jenny
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
/ FILE REFERENCE: PA-0041 US
/ CURRENT APPLICATION NUMBER: US/09/976,594
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: 60/240,409
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 1143
/ SOFTWARE: PERL Program
/ SEQ ID NO 726
/ LENGTH: 2468
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

Query Match          14.2%; Score 92.5; DB 2; Length 2468;
Best Local Similarity 31.0%; Pred. No. 0.83;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKDTFIKPVFKKIEKKEENKPTFDVSKKDDNPVNHSQLNESHKEDLQRE-----E 92
Db 638 VKKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKKEKPKKEEVKKEIK 689
```









Query Match 13.2%; Score 86; DB 2; Length 243;  
Best Local Similarity 28.0%; Pred. No. 0.2;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
  
QY 16 YQFGWEISGFGKDGAGVNLKDTTPKVPFK-----IEKKKEENKPTFDVSKK-- 69  
DB 98 YDDDDDEFGPSSGAAKELNLSQAKEWKQRDLIEIEREKLNSKKKEIEIEAKS 157  
  
QY 70 --DNPQVNHSQLNESHKEDLOREH--SQKSDSTKDVATVLDKVN 112  
DB 158 TIDDPYENYNGKRDNHQKLEILQEQKFISKRDDFLK--RGLWDRVN 202  
  
RESULT 21  
US-09-949-016-10076  
; Sequence 10076, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: PsetSeq for Windows Version 4.0  
; SEQ ID NO 10076  
; LENGTH: 1989  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10076  
  
Query Match 13.2%; Score 86; DB 2; Length 1989;  
Best Local Similarity 26.2%; Pred. No. 3.2;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;  
  
QY 17 OPDGEISGFGKDGAGVNLKDTTPKVPFKIEKKKEENKPTFDVSKKQNPQVNH 76  
DB 1002 EWNQLQISVIRIKKGVAT-CLKVAFMQAHFK---QREAEVKPLDLYEKKANCIAH 1057  
  
QY 77 SOLNESHKEDLOREHSQKSDSTKDVATVLDKNNISKSTNNPN 123  
DB 1058 TGA-DIHRNGDFQKNGTGTSGIGSSVEKYIIDEDHM--SPINPN 1100  
  
RESULT 22  
US-09-248-796A-24668  
; Sequence 24668, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 24668  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-24668

Query Match 13.1%; Score 85.5; DB 2; Length 109;  
Best Local Similarity 32.9%; Pred. No. 0.079;  
Matches 27; Conservative 14; Mismatches 26; Indels 15; Gaps 4;  
  
QY 52 EKKKEEN-----KPTFDVSK--IKDNPOVNHSQLNESHKEDLOREHS-QKSDSTK 101  
DB 11 DDESEETKDKSGKEDIDRNKDSIEDNSNANSTQAVNKLTKLENEHSDKSDPTK 70  
  
QY 102 DVTATVLDKNNISKSTNNPN 123  
DB 71 ENS-----KDGKVSKEENTNAN 87  
  
RESULT 23  
US-09-134-001C-3856  
; Sequence 3856, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3856  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3856  
  
Query Match 13.1%; Score 85; DB 2; Length 465;  
Best Local Similarity 32.6%; Pred. No. 0.61;  
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;  
  
QY 51 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLO-----REEHSQKSDS--- 99  
DB 3 MEENKQPNKE--NMSNKDDNA-----THLNDSHREDELELFRNKNARQRRRRIDNQS 56  
  
QY 100 TKDVTAT-----VLDKNNISKSTNNPNK 124  
DB 57 EKDATSTQSQLETKPMDKPIDNHKS--HNQNK 86  
  
RESULT 24  
US-09-710-279-658  
; Sequence 658, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 658  
; LENGTH: 472  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-658  
  
Query Match 13.1%; Score 85; DB 2; Length 472;



Qy	17	QFGWEISGPEGKDGAGVINLSKDTTFIKPVFKIEBKKEEENKPTFDVSKKONPQVNH	76
		: : :       : : :       : : :       : : :	
Db	989	EMNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKCANCTANH	1044
Qy	77	SQLNESHKREDQREEHSSQKSDSTKVDTATVLDKNNISSKSTNNPN	123

RESULT 31  
US-09-024-020B-3  
; Sequence 3, Application US/09024020B  
; Patent No. 6030810  
; GENERAL INFORMATION:  
; APPLICANT: DELGADO, STEPHEN G.  
;

APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 09/024,020B  
FILING DATE: 16-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,447  
FILING DATE: 26-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, JANET P.  
REGISTRATION NUMBER: 34,799  
REFERENCE/DOCKET NUMBER: R0020B-REG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 852-3097  
TELEFAX: (650) 855-5322  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1978 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-024-020B-3  
Query Match 13.1%; Score 85; DB 2; Length 1978;  
Best Local Similarity 25.2%; Pred. No. 4.1;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
QY 17 QFDGWEISGFEKKDAGYVNLSDTTPKVPFKKIEKKKEENKPTFDVSKKONPQVNH 76  
Db 991 EMNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKCNCIANH 1046  
QY 77 SOLNESHKEDLOREHSQKSDSTKDTATVLDKNNISSKSTNNPN 123  
Db 1047 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1089  
RESULT 32  
US-09-425-043-3  
Sequence 3, Application US/09425043  
Patent No. 6335172  
GENERAL INFORMATION:  
APPLICANT: DELGADO, STEPHEN G.  
APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA

COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/425,043  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/024,020  
FILING DATE: 16-FEB-1998  
APPLICATION NUMBER: US 60/039,447  
FILING DATE: 26-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, JANET P.  
REGISTRATION NUMBER: 34,799  
REFERENCE/DOCKET NUMBER: R0020B-REG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 852-3097  
TELEFAX: (650) 855-5322  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1978 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-425-043-3  
Query Match 13.1%; Score 85; DB 2; Length 1978;  
Best Local Similarity 25.2%; Pred. No. 4.1;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
QY 17 QFDGWEISGFEKKDAGYVNLSDTTPKVPFKKIEKKKEENKPTFDVSKKONPQVNH 76  
Db 991 EMNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKCNCIANH 1046  
QY 77 SOLNESHKEDLOREHSQKSDSTKDTATVLDKNNISSKSTNNPN 123  
Db 1047 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1089  
RESULT 33  
US-09-024-020B-4  
Sequence 4, Application US/09024020B  
Patent No. 6030810  
GENERAL INFORMATION:  
APPLICANT: DELGADO, STEPHEN G.  
APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,020B  
FILING DATE: 16-FEB-1998

```
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/039,447
/ FILING DATE: 26-FEB-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: CLARK, JANET P.
/ REGISTRATION NUMBER: 34,799
/ REFERENCE/DOCKET NUMBER: R0020B-REG
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 852-3097
/ TELEFAX: (650) 855-5322
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1988 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-09-024-020B-4

Query Match 13.1%; Score 85; DB 2; Length 1988;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 17 QFDGWEISGPEGKDGAGYVNLKDTFKIPVFKIEEKEENKPTFDVSKKDNPOVNH 76
Db 1001 EMNLIQISVIRIKGVAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCINAH 1056
Qy 77 SOLNESHKEDLQREHSQKSDSTKVTTATVLDKNNISSKSTTNNPN 123
Db 1057 TGV-DIHRNGDFQKNGTSGIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 34
US-09-425-043-4
Sequence 4, Application US/09425043
Patent No. 6335172
GENERAL INFORMATION:
APPLICANT: DELGADO, STEPHEN G.
APPLICANT: DIETRICH, PAUL S.
APPLICANT: FISH, LINDA M.
APPLICANT: HERMAN, RONALD C.
APPLICANT: SANGAMESWARAN, LAKSHMI
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
TITLE OF INVENTION: SODIUM CHANNEL 1-SUBUNIT AND A SPLICE VARIANT THEREOF
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: JANET PAULINE CLARK
STREET: 3401 HILLVIEW AVENUE, MS A2-250
CITY: PALO ALTO
STATE: CA
COUNTRY: U.S.A.
ZIP: 94304-1397
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/425,043
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/024,020
FILING DATE: 16-FEB-1998
APPLICATION NUMBER: US 60/039,447
FILING DATE: 26-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: CLARK, JANET P.
REGISTRATION NUMBER: 34,799
REFERENCE/DOCKET NUMBER: R0020B-REG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 852-3097
```

```
/ TELEFAX: (650) 855-5322
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1988 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-09-425-043-4

Query Match 13.1%; Score 85; DB 2; Length 1988;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 17 QFDGWEISGPEGKDGAGYVNLKDTFKIPVFKIEEKEENKPTFDVSKKDNPOVNH 76
Db 1001 EMNLIQISVIRIKGVAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCINAH 1056
Qy 77 SOLNESHKEDLQREHSQKSDSTKVTTATVLDKNNISSKSTTNNPN 123
Db 1057 TGV-DIHRNGDFQKNGTSGIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 35
US-10-172-502-10
Sequence 10, Application US/10172502
Patent No. 6841154
GENERAL INFORMATION:
APPLICANT: POSTER, Timothy et al.
TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
FILE REFERENCE: P07263US01/BAS
CURRENT APPLICATION NUMBER: US/10/172,502
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: US 60/298,098
PRIOR FILING DATE: 2001-06-15
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 654
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match 13.0%; Score 84.5; DB 2; Length 654;
Best Local Similarity 28.6%; Pred. No. 1.1;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 31 DAGYVINL-SKDTFKIPVFKIEEKEENKPTFDV-----SKKDNPOVNHSQLNESH 84
Db 450 DGQYHVRIVDKEAFTKANTDKSNKKEQQDNSAKKEATPATPSKPTSPVKESSQKDSQK 509
Qy 85 KEDLQ-----REHSQKSDSTKVDT-ATVLDKNNISSKSTTNNPNK 124
Db 510 DDNKQLPSVEKENDASSESGDKTTPATKPTKGEVSSSTT--PTK 552

RESULT 36
US-09-248-796A-20276
Sequence 20276, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20276
```

; LENGTH: 817

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20276

Query Match

12.9%; Score 84; DB 2; Length 817;

Best Local Similarity

27.0%; Pred. No. 1.6;

Matches 24; Conservative 17; Mismatches 34; Indels 14; Gaps 4;

QY 33 GVINLSK-----DTFKPVPFKIEEKEENKPTFDVSKKQNPQVNHSQLNESHKED 87

Db 312 GLTISQSLDKLASTVQPIILDIGKAEA--KRQIDIEKKQ-----KELEIQLLHEK-- 362

QY 88 LQREHSQKSDTKDVTATVLDKNNISSK 116

Db 363 AKKEEHEAKEKEKRDIEIAKLERNANDK 391

RESULT 37

US-09-949-016-8508

; Sequence 8508, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8508

; LENGTH: 461

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-8508

Query Match

12.8%; Score 83.5; DB 2; Length 461;

Best Local Similarity

21.7%; Pred. No. 0.87;

Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 30 KDAGVINLSKDTFKPVPFKIEEKK-----EENKPTFDVSKKQNPQ 73

Db 26 RDSGLSQBEEDFTFER--QQLSEKJLERERLHEEWLLREQKAQEFKKEKRAA 83

QY 74 VN-----HSQLNESHKEDLQREHSQKSDTKDVTATVLD--KNNISSKSTTNP 122

Db 84 KKWLEEQERKLUKEQWKEQORKEEREEOQKEKKEEAVQKMLDQANDLSTTQNP 143

RESULT 38

US-09-248-796A-15008

; Sequence 15008, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 15008

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-15008

Query Match

12.8%; Score 83.5; DB 2; Length 476;

Best Local Similarity

27.6%; Pred. No. 0.91;

Matches 27; Conservative 19; Mismatches 27; Indels 25; Gaps 3;

QY 38 LSKDTFKPVPFKIEEKEENKPTFDVSKKQNPQVNH-----SQL 79

Db 181 LKDWKAK---LKLEKQKQNDLTKDLFKKQDTPASNPFGGNSNPFGLNPFSEK 237

QY 80 NESHKEDLQREHSQKSDTKDVTATVLDKNNISSKS 117

Db 238 PEEKEBEKEKEKETSKEYADVAS----KNAPKPKS 271

RESULT 39

US-09-949-016-10508

; Sequence 10508, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10508

; LENGTH: 278

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-10508

Query Match

12.7%; Score 83; DB 2; Length 278;

Best Local Similarity

25.4%; Pred. No. 0.51;

Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;

QY 8 YKGELEKGYQDGEHISGPEGKKDAGY--VINLSKDTFKPVPFKIEEKEENKPTFDV 65

Db 59 WEGSEDEVDKNDWDDDDDEKKEBAFVKPEVKISEK---KKIAEKIKKEKQKQKQBEI 115

QY 66 SKKQNPQ-----VNHSQLNESHKEDLQREHSQKSDTKDVTATV--LDKNNISSK 116

Db 116 KRLREPEEPKULTPESQLADKLKLKQLQESDLELAKETFGVNNVATGIDAMPSSR 173

RESULT 40

US-09-248-796A-21334

; Sequence 21334, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 21334



```
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21334

Query Match      12.7%; Score 82.5; DB 2; Length 243;
Best Local Similarity 30.8%; Pred. No. 0.48;
Matches 24; Conservative 16; Mismatches 27; Indels 11; Gaps 4;

QY 52 EKKEEENKPTDVS---KKQDNPOVNHSQLNESHKEDLQREEHSQLKSDSTKDVATVL 108
   ||::: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 71 EEEQQQQQIEDVSGYISPFDPNPIHSGK----HKRHLKND-SISNSSNKEITIDAI 125
   ||::: : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :

QY 109 DK--NNISSKSTTNP 123
   ||| ||| ||| |||
Db 126 IKRQNNISSNDNNNNN 143
   ||| ||| ||| |||

Search completed: April 24, 2006, 15:03:35
Job time : 21.2796 secs
```

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 77.8121 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDPLPVYKGELEKQYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	4	AAB48343
2	651	100.0	2120	3	AA81710 Streptococcus
3	651	100.0	2140	6	ABU01020 S. pneumo
4	651	100.0	2140	6	ABU45746 Protein e
5	651	100.0	2140	8	ADM92113 S. pneumo
6	651	100.0	2140	8	ADT50099 S. pneumo
7	648	99.5	637	8	ADR94534 Novel S.
8	648	99.5	637	9	AEA58404 Streptococcus
9	648	99.5	2138	8	ADK48759 Streptococcus
10	615	94.5	117	2	AAW55096 Streptococcus
11	615	94.5	117	5	ABP54590 Streptococcus
12	615	94.5	117	7	ADC45149 Streptococcus
13	106	16.3	188	9	ADZ79639 P. falcip
14	106	16.3	354	9	ADZ72253 Plasmodium
15	101.5	15.6	564	4	ABB61977 Drosophila
16	101	15.5	169	9	ADZ79634 P. falcip
17	101	15.5	647	9	ADZ79635 P. falcip
18	101	15.5	651	8	ADO19012 Amino aci
19	101	15.5	651	8	ADO19010 P. falcip
20	98	15.1	665	3	AA818278 Streptococcus
21	98	15.1	665	7	ABO23606 Plasmodium
22	96.5	14.8	707	6	ABU25018 Protein e
23	92.5	14.2	2468	6	ABR64281 Angiogene
24	92.5	14.2	2468	7	AD662723 Human Pro

25	92.5	14.2	2468	7	AD662719 Human Pro
26	92.5	14.2	2468	7	AD662727 Human Pro
27	92.5	14.2	2468	7	AD662715 Human Pro
28	92.5	14.2	2468	8	ADL12997 Human ste
29	92.5	14.2	2468	8	ADN05260 Antipsori
30	92.5	14.2	2468	8	ADR14614 Human NF-
31	92.5	14.2	2519	4	ASG16636 Novel hum
32	92.5	14.2	2527	8	ADN04561 Antipsori
33	91	14.0	470	8	ADT56185 Plant pol
34	91	14.0	484	3	AA847777 Arabidops
35	90.5	13.9	639	9	ADW88474 Staphyloc
36	89.5	13.7	511	2	AA835091 Chlamydia
37	89.5	13.7	645	9	ADW88441 Staphyloc
38	88	13.5	225	5	ABP73992 Candida a
39	88	13.5	258	4	AA894584 Human pro
40	88	13.5	815	7	ADP28113 Murine gl
41	88	13.5	817	4	AAW79318 Human pro
42	88	13.5	817	4	AAW79319 Human pro
43	88	13.5	1980	7	ADB78600 Human bod
44	87.5	13.4	903	6	ABU24404 Protein e
45	87.5	13.4	1702	6	ADA09346 Haemophil

#### ALIGNMENTS

##### RESULT 1

AA848343  
ID AAB48343 standard; protein; 773 AA.

XX  
AC AAB48343;

XX  
DT 20-APR-2001 (first entry)

XX  
DE S. pneumoniae Spl30 polypeptide.

XX  
KW Immunogenic; Spl28; Spl30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.

OS Streptococcus pneumoniae.

XX  
FN WO200076540-A2.

XX  
PD 21-DEC-2000.

XX  
PF 09-JUN-2000; 2000WO-US015925.

XX  
PR 10-JUN-1999; 99US-0138453P.

XX  
PA (MEDI-) MED IMMUNE INC.

XX  
PI Adamou JE, Choi GH;

XX  
XX WPI; 2001-112197/12.

DR N-PSDB; AAC84742.

XX  
PT New vaccines comprising Spl28 or Spl30 polypeptides, for treating and  
PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections.

PS Claim 8; Page 51-54; 54pp; English.

XX  
CC The invention relates to novel immunogenic polypeptides, Spl28 and Spl30  
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.

CC  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spl30 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 651; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 1.3e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKKEENK 60  
| | | | |  
DB 650 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKKEENK 709  
| | | | |

QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTN 120  
| | | | |  
DB 710 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTN 769  
| | | | |

QY 121 NPNK 124  
| | | | |  
DB 770 NPNK 773  
| | | | |

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PP 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 651; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 5e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKKEENK 60  
| | | | |  
DB 1963 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKKEENK 2022  
| | | | |

QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTN 120  
| | | | |  
DB 2023 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTN 2082  
| | | | |

QY 121 NPNK 124  
| | | | |  
DB 2083 NPNK 2086  
| | | | |

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PP 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC ABS56454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence, and  
 CC the second primer is substantially complementary to the target sequence of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 651; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 2042  
 QY 61 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.  
 XX  
 AC ABU45746;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #31273.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 73670; 1766pp; English.  
 XX

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 651; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 2042  
 QY 61 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.  
 XX  
 AC ADM92113;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE S pneumoniae antigenic protein sequence SeqID310.  
 XX  
 KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 KW antigenic.  
 XX  
 OS Streptococcus pneumoniae.

PN WO2004020609-A2.  
 XX 11-MAR-2004.  
 XX PF 02-SEP-2003; 2003WO-US027401.  
 XX PR 30-AUG-2002; 2002US-0407082P.  
 XX PA (TUFT ) UNIV TUFTS.  
 XX PI Camilli A, Hava DL;  
 XX WPI; 2004-239189/22.  
 DR N-PSDB; ADM91876.  
 XX  
 PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
 PT diagnosing, treating and preventing active infections of Streptococcus  
 PT pneumoniae.  
 XX  
 PS Claim 27; SEQ ID NO 310; 123pp; English.  
 CC  
 CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
 CC acid molecules and the antigenic polypeptides encoded by them. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
 CC compositions and methods disclosed are useful for treating Streptococcus  
 CC pneumoniae infection. The present sequence is that of an S pneumoniae  
 CC protein of the invention.  
 XX  
 XX Sequence 2140 AA;  
 SQ  
 Query Match 100.0%; Score 651; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 2042  
 QY 61 PTFDVSKKDNQPNVHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNQPNVHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106  
 RESULT 6  
 ADT50099  
 ID ADT50099 standard; protein; 2140 AA.  
 XX  
 XX AC ADT50099;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
 XX  
 KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
 KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
 KW sepsis; meningitis.  
 XX  
 OS Streptococcus pneumoniae TIGR4.  
 XX  
 PN WO2004092209-A2.  
 XX  
 XX PD 28-OCT-2004.  
 XX  
 XX PF 15-APR-2004; 2004WO-EP003984.  
 XX PR 15-APR-2003; 2003EP-00450087.  
 XX PA (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
 XX WPI; 2004-758335/74.  
 DR N-PSDB; ADT49955.  
 XX  
 PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
 PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
 PT treating S. pneumoniae infections.  
 XX  
 PS Disclosure; SEQ ID NO 177; 191pp; English.  
 CC  
 CC This invention relates to novel nucleic acids encoding hyperimmune serum  
 CC reactive antigens, or fragments derived thereof. Specifically, it refers  
 CC to antigens selected from peptides and serum reactive epitopes that can  
 CC be used in pharmaceutical compositions that exhibit antibacterial  
 CC activity. The present invention describes a composition (including the  
 CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
 CC that is useful for manufacturing a medicament such as a vaccine, which  
 CC can be used to treat or prevent bacterial infections, particularly S.  
 CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
 CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
 CC be used for isolating, purifying and/ or identifying an interaction  
 CC partner of the hyperimmune serum reactive antigen, as well as for  
 CC manufacturing a functional nucleic acid selected from aptamers and  
 CC Spiegelmers or for manufacturing a functional ribonucleic acid selected  
 CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
 CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
 CC of the invention.  
 XX  
 XX Sequence 2140 AA;  
 SQ  
 Query Match 100.0%; Score 651; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 2042  
 QY 61 PTFDVSKKDNQPNVHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNQPNVHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106  
 RESULT 7  
 ADT94534  
 ID ADT94534 standard; protein; 637 AA.  
 XX  
 XX AC ADT94534;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
 XX  
 KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US6800744-B1.  
 XX  
 XX PD 05-OCT-2004.  
 XX  
 XX PF 30-JUN-1998; 98US-00107433.  
 XX PR 02-JUL-1997; 97US-0051553P.  
 XX PR 12-MAY-1998; 98US-0085131P.  
 XX

PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Doucette-Stamm LA, Bush D;  
XX  
XX  
DR WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
XX  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX  
PS Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
XX  
SQ Sequence 637 AA;  
Query Match 99.5%; Score 648; DB 8; Length 637;  
Best Local Similarity 99.2%; Pred. No. 2e-58;  
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EDFILPVYKGELEKGYQFDGWEISGFEGKDGAGVYVNLSDTFIKPVPKKEEKEENK 60  
Db 480 EDFILPVYKGELEKGYQFDGWEISGFEGKDGAGVYVNLSDTFIKPVPKKEEKEENK 539  
QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
Db 540 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 599  
QY 121 NPNK 124  
Db 600 NPNK 603  
RESULT 8  
AEA58404  
ID AEA58404 standard; protein; 637 AA.  
XX  
XX AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
XX  
XX Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
XX  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
KW vaccine.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX US2005136404-A1.  
XX

PD 23-JUN-2005.  
XX  
XX 10-JUL-2003; 2003US-00617320.  
XX  
XX 02-JUL-1997; 97US-0051553P.  
XX 12-MAY-1998; 98US-0085131P.  
XX 30-JUN-1998; 98US-00107433.  
XX  
XX (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
XX New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
XX Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
XX  
SQ Sequence 637 AA;  
Query Match 99.5%; Score 648; DB 9; Length 637;  
Best Local Similarity 99.2%; Pred. No. 2e-58;  
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EDFILPVYKGELEKGYQFDGWEISGFEGKDGAGVYVNLSDTFIKPVPKKEEKEENK 60  
Db 480 EDFILPVYKGELEKGYQFDGWEISGFEGKDGAGVYVNLSDTFIKPVPKKEEKEENK 539  
QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 120  
Db 540 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 599  
QY 121 NPNK 124  
Db 600 NPNK 603  
RESULT 9  
ADK48759  
ID ADK48759 standard; protein; 2138 AA.

```
XX AC ADK48759;
XX DT 20-MAY-2004 (first entry)
XX DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX OS Streptococcus pneumoniae.
XX PN US6699703-B1.
XX PD 02-MAR-2004.
XX PF 26-MAY-2000; 2000US-00583110.
XX PR 02-JUL-1997; 97US-0051553P.
XX PR 12-MAY-1998; 98US-0085131P.
XX PR 30-JUN-1998; 98US-00107433.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX DR N-PSDB; ADK46098.
XX DE New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX SQ Sequence 2138 AA;
Query Match 99.5%; Score 648; DB 8; Length 2138;
Best Local Similarity 99.2%; Pred. No. 1e-57;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDFILPVYKGELEKGYQPDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEENK 60
DB 1981 EDFILPVYKGELEKGYQPDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEENK 2040
QY 61 PTFDYSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120
DB 2041 PTFDYSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 2100
QY 121 NFNK 124
DB 2101 NFNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX AC AAW55096;
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0043 protein.
```

```
XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX OS detection; pneumonia; otitis media; meningitis.
XX PN Streptococcus pneumoniae.
XX PD WO9818930-A2.
XX PF 07-MAY-1998.
XX PR 30-OCT-1997; 97WO-US019422.
XX PR 31-OCT-1996; 96US-0029960P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX DR WPI; 1998-272224/24.
XX DR N-PSDB; AAV27357.
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX PS Claim 11; Page 62; 118pp; English.
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose
XX SQ Sequence 117 AA;
Query Match 94.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.8e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 YKGELEKGYQPDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEENKPTFDYSK 67
DB 1 YKGELEKGYQPDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEENKPTFDYSK 60
QY 68 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 124
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX AC ABP54590;
XX DT 04-SEP-2002 (first entry)
XX DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX KW antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
XX PN US2002061545-A1.
```



XX PD 23-MAY-2002.  
 XX PF 22-JAN-2001; 2001US-00765272.  
 XX PR 30-OCT-1997; 97US-00961083.  
 XX PA (CHOI/) CHOI G H.  
 XX PA (KUNS/) KUNSCH C A.  
 XX PA (BARA/) BARASH S C.  
 XX PA (DILL/) DILLON P J.  
 XX PA (DOUG/) DOUGHERTY B.  
 XX PA (FANN/) FANNON M R.  
 XX PA (ROSE/) ROSEN C A.  
 XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 XX PI Rosen CA;  
 XX DR WPI; 2002-479261/51.  
 XX DR N-PSDB; ABQ84825.  
 XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
 XX PT and for preventing or attenuating disease caused by Streptococcus  
 XX PT infection.  
 XX PS Claim 11; Page 29; 70pp; English.  
 XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
 XX CC pneumoniae antigens have antibacterial activity and can be used in  
 XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
 XX CC attenuate a Streptococcal infection in an animal. The polynucleotides  
 XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
 XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning  
 XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
 XX CC from the present invention  
 XX SQ Sequence 117 AA;  
 Query Match 94.5%; Score 615; DB 5; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-56;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSK 67  
 Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSK 60  
 QY 68 KKDNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 124  
 Db 61 KKDNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117  
 RESULT 12  
 ADC45149  
 ID ADC45149 standard; protein; 117 AA.  
 XX AC ADC45149;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE S. pneumoniae antigenic protein SP043.  
 XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.  
 XX OS Streptococcus pneumoniae.  
 XX PN US6573082-B1.  
 XX PD 03-JUN-2003.  
 XX PF 28-MAR-2000; 2000US-00536784.  
 XX PR 31-OCT-1996; 96US-0029960P.

PR 30-OCT-1997; 97US-00961083.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 XX PI Rosen CA;  
 XX DR WPI; 2003-764574/72.  
 XX DR N-PSDB; ADC45148.  
 XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
 XX PT useful for producing vaccines for prevention or attenuation of infection  
 XX PT by Streptococcus pneumoniae.  
 XX PS Example 1; SEQ ID NO 68; 58pp; English.  
 XX CC The invention relates to an isolated polynucleotide consisting of a  
 XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
 XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
 XX CC antigens. Also included are making a recombinant vector by inserting the  
 XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
 XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
 XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
 XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
 XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
 XX CC antigen nucleic acids are useful as probes for use in diagnostic methods  
 XX CC for detecting S. pneumoniae gene expression. The present sequence  
 XX CC represents an S. pneumoniae antigenic protein.  
 XX SQ Sequence 117 AA;  
 Query Match 94.5%; Score 615; DB 7; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-56;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSK 67  
 Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSK 60  
 QY 68 KKDNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 124  
 Db 61 KKDNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117  
 RESULT 13  
 ADZ79639  
 ID ADZ79639 standard; protein; 188 AA.  
 XX AC ADZ79639;  
 XX DT 14-JUL-2005 (first entry)  
 XX DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.  
 XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 XX KW immunotherapy; malaria; antimalarial; vaccine.  
 XX OS Plasmodium falciparum.  
 XX PN WO2005040206-A1.  
 XX PD 06-MAY-2005.  
 XX PF 22-OCT-2004; 2004WO-EP012910.  
 XX PR 24-OCT-2003; 2003US-00691672.  
 XX PA (INSP) INST PASTEUR.  
 XX PI Druilhe P;  
 XX DR WPI; 2005-355821/36.

PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

XX  
 PS Disclosure; SEQ ID NO 7; 79pp; English.

XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
 CC MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the  
 CC Sequence Listing is not mentioned elsewhere in the specification.

XX  
 SQ Sequence 188 AA;

Query Match 16.3%; Score 106; DB 9; Length 188;  
 Best Local Similarity 23.5%; Pred. No. 0.012;  
 Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;  
 QY 7 VYKGELEKGYQFD--GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEE- 57  
 Db 15 VLKAKEASSYDYLGWEGFGVPEHKKEENMLSHLYVSSKDKENISKENDDVLDKEBEEA 74  
 QY 58 -----ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDYTA 105  
 Db 75 ETEEEELKEKNEEETESISEDEEEEBEKEEENDKKKEQKEQSNENNDDQKDMA 134  
 QY 106 TVLDKNNISSKSTTN 121  
 Db 135 -----QNLISKQNNN 145

RESULT 14

ADZ72253  
 ID ADZ72253 standard; protein; 354 AA.

XX  
 AC ADZ72253;

XX  
 DT 14-JUL-2005 (first entry)

XX  
 DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.

XX  
 KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
 KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.

XX  
 OS Plasmodium falciparum.

XX  
 PN EP1526178-A1.

XX  
 PD 27-APR-2005.

XX  
 PF 24-OCT-2003; 2003EP-00292673.

XX  
 PR 24-OCT-2003; 2003EP-00292673.

XX  
 PA (INSP ) INST PASTEUR.

XX

PI Druilhe P;

XX  
 DR WPI; 2005-322987/34.

XX  
 DR N-PSDB; ADZ72252.

XX  
 PT Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
 PT falciparum, which encode proteins useful for preparing vaccine  
 PT compositions against malaria.

XX  
 PS Disclosure; SEQ ID NO 2; 137pp; English.

XX  
 CC The present invention relates to the protection against malaria. More  
 CC particularly, the invention pertains to a family of MSP-3 (merozoite  
 CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
 CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
 CC falciparum, highly conserved in P. falciparum strains, simultaneously  
 CC expressed in P. falciparum at the erythrocytic stages and encoding  
 CC proteins which have a Asn-Leu-Arg-Hen or Asn-Leu-Arg-Lys signature at  
 CC their N-terminal extremity and which are located at the merozoite  
 CC surface. The characterization of this gene family enables the definition  
 CC of immunogenic and vaccine compositions against P. falciparum. The  
 CC present sequence is the P. falciparum MSP-3-1 protein.

XX  
 SQ Sequence 354 AA;

Query Match 16.3%; Score 106; DB 9; Length 354;  
 Best Local Similarity 23.5%; Pred. No. 0.029;  
 Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;  
 QY 7 VYKGELEKGYQFD--GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEE- 57  
 Db 181 VLKAKEASSYDYLGWEGFGVPEHKKEENMLSHLYVSSKDKENISKENDDVLDKEBEEA 240  
 QY 58 -----ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDYTA 105  
 Db 241 ETEEEELKEKNEEETESISEDEEEEBEKEEENDKKKEQKEQSNENNDDQKDMA 300  
 QY 106 TVLDKNNISSKSTTN 121  
 Db 301 -----QNLISKQNNN 311

RESULT 15

ABB61977

ID ABB61977 standard; protein; 564 AA.

XX  
 AC ABB61977;

XX  
 DT 26-MAR-2002 (first entry)

XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 12723.

XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX  
 OS Drosophila melanogaster.

XX  
 PN WO200171042-A2.

XX  
 PD 27-SEP-2001.

XX  
 PF 23-MAR-2001; 2001WO-US009231.

XX  
 PR 23-MAR-2000; 2000US-0191637P.

XX  
 PR 11-JUL-2000; 2000US-00614150.

XX  
 PA (PEKE ) PE CORP NY.

XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
 DR WPI; 2001-656860/75.

XX  
 DR N-PSDB; ABL06080.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

PS Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 564 AA;

Query Match 15.6%; Score 101.5; DB 4; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.16;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 1 EDFILPVYKGELEKGVQDGM-----ETSGPEGKKDAGYVI-----NLSKDTFIK 45  
DB 78 ELDLTPLESRFSK--VFDGWVDEHRDHDGHVQSPGSEALDDHDEHDDHDDHDEDEE 135  
QY 46 PVFKKIKKKEENKPT-----FDVSKKNQPNVNSQLNESHKEDLQREHHSQKSDS 99  
DB 136 PLTELELELELEEEPTEDRPAADVEYEDEDEENNA--GENTTAEDAEDEEEDND 193

QY 100 TKDVTATVLDKNNISSKST 118  
DB 194 EGTVEATVEATTEAT 212

RESULT 16  
ADZ79634  
ID ADZ79634 standard; protein; 169 AA.  
XX AC ADZ79634;  
XX DT 14-JUL-2005 (first entry)  
XX DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.  
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
XX KW immunotherapy; malaria; antimalarial; vaccine.  
XX OS Plasmodium falciparum.  
XX FH Key  
XX FT Location/Qualifiers  
XX FT Region  
XX FT /note= "Amino acid residues 212-380 of MSP3"  
XX FN WO2005040206-A1.  
XX PD 06-MAY-2005.  
XX PP 22-OCT-2004; 2004WO-EP012910.  
XX XX 24-OCT-2003; 2003US-00691672.  
XX XX (INSP ) INST PASTEUR.  
XX XX Druilhe P;  
XX XX WPI; 2005-355821/36.  
XX PT Chimeric molecule useful for preparing vaccine composition against  
XX PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
XX PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
XX PT immunized with molecule.

XX Claim 2; SEQ ID NO 2; 79pp; English.

XX The invention relates to a chimeric molecule that comprises a glutamate-  
XX rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
XX acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
XX surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
XX of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
XX antibodies against both polypeptides in mice immunized with it. Also  
XX described are: (i) a conjugate comprising the chimeric molecule of the  
XX invention bound to a solid support, (ii) an immunogenic composition  
XX comprising the chimeric molecule, the conjugate described above, or a  
XX mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
XX against malaria comprising the chimeric molecule, the conjugate described  
XX above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
XX association with a suitable vehicle, (iv) use of purified and/or  
XX recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
XX medicament against malaria, and (v) a medicament for passive  
XX immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
XX antibodies. The chimeric molecule of the invention or a mixture of GLURP  
XX and MSP3 antigens are useful for the preparation of a vaccine composition  
XX against malaria. This sequence represents Plasmodium falciparum MSP3  
XX protein (amino acid residues 212-380).

XX Sequence 169 AA;

Query Match 15.5%; Score 101; DB 9; Length 169;  
Best Local Similarity 25.4%; Pred. No. 0.036;  
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

QY 20 GWEISGP--EQKQAG-----YVINLSKDTPIKPVFKIKKKEENKPTFDVSKKNP 72  
DB 11 GWEPGGVPEHKKEENMLSHLYVSSKKNISKENDVDLDE--KEEAEETEELEEKNE 69  
QY 73 QVNHSQNLN-----ESHRKEDLQREHHSQKSDSTKVATVLDKNNISSKSTT 119  
DB 70 EETSEISEDEEEEEEKKEEENKKEKQKQENNDNQKQMEA-----QNLISKQN 124

QY 120 NN 121  
DB 125 NN 126

RESULT 17  
ADZ79635  
ID ADZ79635 standard; protein; 647 AA.  
XX AC ADZ79635;  
XX DT 14-JUL-2005 (first entry)  
XX XX P. falciparum GLURP-MSP3 fusion protein.  
XX DE immune stimulation; fusion protein; glutamate-rich protein; GLURP;  
XX KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;  
XX KW vaccine.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX XX WO2005040206-A1.  
XX XX 06-MAY-2005.  
XX XX 22-OCT-2004; 2004WO-EP012910.  
XX XX 24-OCT-2003; 2003US-00691672.  
XX XX (INSP ) INST PASTEUR.  
XX XX Druilhe P;  
XX XX WPI; 2005-355821/36.





PI Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;  
XX WPI; 2003-492159/46.  
XX  
XX Identifying candidate proteins useful as anti-infectives involves  
PT matching outlier protein sequences with protein sequences in databases.  
XX  
XX Example 7; Page 91-93; 117pp; English.  
XX  
XX The present invention relates to a method for identifying candidate  
CC proteins in pathogens useful as anti-infectives. The invention discloses  
CC a computational method which involves the calculation of several sequence  
CC attributes and their subsequent analysis results in the identification  
CC of outlier proteins in different pathogens. The method is useful for the  
CC identification of outlier proteins (e.g. virulence proteins, antigens or  
CC proteins used as drug targets) in pathogenic organisms. The method on the  
CC invention provides reproducible results as it does not depend on the  
CC variable biochemical characterisation of proteins. ABO23500-ABO23617  
CC represent outlier proteins identified from different pathogenic organisms  
XX  
XX Sequence 665 AA;

Query Match 15.1%; Score 98; DB 7; Length 665;  
Best Local Similarity 27.1%; Pred. No. 0.46;  
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;  
QY 13 EKGQFDGWEI--SGFEGKDGAGYVNLKOTFIKPVFKIEEKE-----EENK 60  
DB 158 EKGKQ----DISNSAENKCD-----VKEGVKELEEKKEKESIDHHKVEENK 201  
QY 61 PTFD----VSKKKNQPNVHNSQLNESHKEDLQR--EEHSQKSDSPKDVTTATVLDKNNISS 115  
DB 202 KSDDHKVEENKSDDHKVEENKSDDHKIEEVKVEEHEEDEE-----DKKEKS 252  
QY 116 KSTNNPNK 124  
DB 253 ENKNKDENK 261

RESULT 22  
ABU25018  
ID ABU25018 standard; protein; 707 AA.  
XX  
AC ABU25018;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by Prokaryotic essential gene #10545.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Clostridium difficile.  
XX  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
XX 06-SEP-2001; 2001US-00948993.  
XX  
XX 25-OCT-2001; 2001US-0342923P.  
XX  
XX 08-FEB-2002; 2002US-0007285A.  
XX  
XX 06-FEB-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyeckind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX N-PSDB; ACA28888.

PT New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 52942; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 707 AA;

Query Match 14.8%; Score 96.5; DB 6; Length 707;  
Best Local Similarity 25.2%; Pred. No. 0.72;  
Matches 34; Conservative 25; Mismatches 53; Indels 23; Gaps 4;  
QY 9 KGELEKGYQDGEWISGFEGKDGAGYVNLKOTFIKPVFKIEEKE-----54  
DB 495 KEEIDSNQDIDGVV---EDKOTTDKEYDSNKEDIIEPENKSKKKKAKLFGFIKKDNEEV 551  
QY 55 -KEEEN---KPTFDVSKKKNQPNVHNSQLNESHKEDLQR--EEHSQKSDSPKDVTTATVLD 109  
DB 552 EQEENLNDISPDILDKPVENNQVKSEETEQNELKE-IKKEEPSQHEEERSVKIEKPI 610  
QY 110 KNNISSKSTNNPNK 124  
DB 611 NNNLDEKVSSNNESK 625  
RESULT 23  
ABR64281  
ID ABR64281 standard; protein; 2468 AA.  
XX  
AC ABR64281;  
XX  
XX 15-OCT-2003 (first entry)  
XX  
XX Angiogenesis protein BNO382.  
XX  
XX Cystostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
XX antipeptidic; antiarteriosclerotic; cardiac; vasotropic; angiogenesis;  
XX gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
XX diabetic retinopathy; cardiovascular disease; atherosclerosis;  
XX ischemic limb disease; coronary artery disease.  
XX

```
OS Homo sapiens.
XX WO2003027285-A1.
XX 03-APR-2003.
XX 19-SEP-2002; 2002WO-AU001282.
XX 27-SEP-2001; 2001AU-00007973.
XX 27-SEP-2001; 2001AU-00007974.
XX 11-OCT-2001; 2001AU-00008210.
XX 29-OCT-2001; 2001AU-00008532.
XX 13-NOV-2001; 2001AU-00008838.
XX 28-AUG-2002; 2002AU-00951032.
XX (BION-) BIONOMICS LTD.
XX Gamble JR, Hahn CN, Vadas MA;
XX WPI; 2003-354655/33.
XX N-PSDB; ACP34559.
XX New angiogenic genes and polypeptides, useful for diagnosing,
PT prognosticating or treating an angiogenesis-related disorder, e.g.
PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or
PT cardiovascular diseases.
XX Claim 15; SEQ ID NO 216; 90pp; English.
XX The invention relates to the isolation of novel genes (ACF34446-ACF34559)
CC encoding proteins (ABR64180-ABR64281) involved in the processes of
CC angiogenesis. The nucleic acid molecules are useful in identifying and/or
CC obtaining full-length human genes involved in an angiogenic process. The
CC nucleic acid molecule, polypeptides or complexes encoded, cells or
CC genetically modified non-human animals derived from these are useful for
CC the screening of candidate pharmaceutical compounds used in treating
CC angiogenesis-related disorders. They are also useful for diagnosing,
CC prognosticating or treating an angiogenesis-related disorder, which
CC involves uncontrolled or enhanced angiogenesis or is a disorder in which
CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,
CC diabetic retinopathy, psoriasis or cardiovascular diseases such as
CC atherosclerosis), or involves inappropriately arrested or decreased
CC angiogenesis or is a disorder in which an expanding vasculature is of
CC benefit (e.g. ischemic limb disease or coronary artery disease). The
CC modulator of expression or activity of the polypeptide encoded by the
CC nucleic acid sequence is useful for manufacturing a medicament for the
CC treatment of an angiogenesis-related disorder. This sequence corresponds
CC to one of the novel angiogenic protein
XX Sequence 2468 AA;
Query Match 14.2%; Score 92.5; DB 6; Length 2468;
Best Local Similarity 31.0%; Pred. No. 10;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
QY 38 LSKDTFFKVPFKKIEKKKEENKPTFDVSKKONQVNHSQLNSHRKEDLQRE-----E 92
Db 638 VKKETKVKP-----EDKKKEKKPKKVKAKEDTKPI---KKKEPKKEVKKEVKKEIK 689
QY 93 HSQKSDSTKOV 103
Db 690 KEEKKEPKKEV 700
RESULT 24
ADE62723
ID ADE62723 standard; protein; 2468 AA.
XX ADE62723;
XX 29-JAN-2004 (first entry)
XX Human Protein NP_005900, SEQ ID NO 8656.
```

```
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX (GEHO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; NP_005900.
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2468 AA;
Query Match 14.2%; Score 92.5; DB 7; Length 2468;
Best Local Similarity 31.0%; Pred. No. 10;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
QY 38 LSKDTFFKVPFKKIEKKKEENKPTFDVSKKONQVNHSQLNSHRKEDLQRE-----E 92
Db 638 VKKETKVKP-----EDKKKEKKPKKVKAKEDTKPI---KKKEPKKEVKKEVKKEIK 689
QY 93 HSQKSDSTKOV 103
Db 690 KEEKKEPKKEV 700
```













PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131444P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.

[illegible]

RESULT 35	
ADW#8474	
ID	ADW88474 standard; protein; 639 AA.
XX	
XX	ADW88474;
XX	
XX	21-APR-2005 (first entry)
DT	
XX	
DE	Staphylococcus aureus hybrid ORF0657n polypeptide.
XX	
XX	ORF0657n; vaccine; antibacterial; protein engineering;
KW	
XX	Staphylococcus aureus infection; muten.
KX	
XX	
OS	Staphylococcus aureus.
OS	Synthetic.
XX	
PN	WO2005009378-A2.
XX	
PD	03-FEB-2005.
XX	
PF	22-JUL-2004; 2004WO-US023522.
XX	
PR	24-JUL-2003; 2003US-0489840P.
XX	
PA	(MERI ) MERCK & CO INC.
XX	
PI	Anderson AS, Kuklin N, Jansen KU;
XX	
XX	WPI; 2005-123069/13.
DR	
XX	
PT	Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,
PT	useful for inducing protective immune response in humans against

```

PT Staphylococcus aureus infection.
PS Claim 7; SEQ ID NO 43; 84pp; English.
XX
XX
CC The present sequence is that of a staphylococcus aureus protein ORF0657n
CC hybrid polypeptide. This is an example of claimed hybrid polypeptide
CC immunogens of the invention ADW8439-ADW88474 that comprise a modified S.
CC aureus ORF0657n sequence ADW8433-ADW88438 containing amino acid
CC substitutions that increase sequence similarity to ORF0190 ADW88432. The
CC hybrid polypeptides contain one or more epitopes for ORF0657n and
CC ORF0190. They were designed by taking into account the similarity and
CC differences between native ORF0657n and ORF0190 protein sequences. The
CC invention also provides nucleic acids encoding these hybrid polypeptides,
CC and a method for evaluating the ability of an immunogen to produce a
CC protective immune response against Staphylococcus infection using an
CC animal (mouse or rat) model. The hybrid polypeptides having therapeutic
CC and diagnostic applications, such as being used to provide protective
CC immunity against S. aureus infection, being used to generate antibodies
CC to detect the presence of S. aureus, and being used to generate
CC therapeutic antibodies that target S. aureus.
XX
SQ Sequence 639 AA;

Query Match 13.9%; Score 90.5; DB 9; Length 639;
Best Local Similarity 29.1%; Pred. No. 2.7;
Matches 32; Conservative 20; Mismatches 43; Indels 15; Gaps 5;

QY 25 GFGEKKDAGYVINLSKDTFTKPVFKTEEEKKEENKPTPDV-----SKCKDNQVNHSQL 79
Db 440 GYEGQY---HVRIVDKFAFTKANTDKENKKEQQDNSAKGKATPATPSKPTSPVEKESQK 496
QY 80 NESHKEDLQ-----REEHSQKSDSTKDVATVLDDKNNISSKSTTNPNK 124
Db 497 QDSQKDNKQLPSVEKENDASSGSKDKTATPKTKGEVSSSTT--PTK 544

RESULT 36
ID AAV35091
XX AAV35091 standard; protein; 511 AA.
AC AAV35091;
XX
XX 17-OCT-2003 (revised)
DT 13-SEP-1999 (first entry)
XX
XX Chlamydia pneumoniae transmembrane protein sequence.
DE
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;
KW neutralising epitope.
XX
XX Chlamydophila pneumoniae.
OS
XX
XX W09927105-A2.
FN
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-IB001890.
PP
XX 21-NOV-1997; 97FR-00014673.
PR 04-NOV-1998; 98US-0107078P.
XX
XX (GEST ) GENSET.
PA
XX
XX Griffais R;
PI
XX WPI; 1999-357842/30.
DR
XX
XX Genome sequence of Chlamydia pneumoniae.
PT
XX Page 975-976; Disclosure; 1912pp; English.
PS
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC

```



CC frames in the complete genome (see AAX91990) of *Chlamydia pneumoniae*. C.  
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis  
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,  
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The  
 CC polypeptides encoded by the open reading frames of the *C. pneumoniae*  
 CC genome (see AAX34584-Y35879) can be used in immunogenic compositions as  
 CC vaccines. Vectors containing *C. pneumoniae* nucleotide sequences can also  
 CC be used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of *C. pneumoniae*. (Updated on 17  
 CC -OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 511 AA;

Query Match 13.7%; Score 89.5; DB 2; Length 511;  
 Best Local Similarity 24.5%; Pred. No. 2.5;  
 Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
 QY 44 IKPVFKTKEEKEENKPTFD-----VSKKDNQPNVHNSQLNESHK 85  
 DB 95 VGVFKTQAPPEVSSPLPSHVHGOQLPGLGFRDRIRKRSNPADLGMKRSYSD 154  
 QY 86 EDLQREHSQKSDTKDTATVLDKNNISSKSTT 119  
 DB 155 GDLDRVGHDSNEDSTEDSR-----EGGEPSSKSS 185

RESULT 37  
 ADW88441  
 ID ADW88441 standard; protein; 645 AA.  
 AC ADW88441;  
 DT 21-APR-2005 (first entry)  
 XX Staphylococcus aureus hybrid polypeptide 0657nHybrid3.  
 DE ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 KW Staphylococcus aureus.  
 OS Synthetic.  
 OS  
 XX WO2005009378-A2.  
 XX  
 XX PD 03-FEB-2005.  
 XX PF 22-JUL-2004; 2004WO-US023522.  
 XX PR 24-JUL-2003; 2003US-0489840P.  
 XX PA (MERI ) MERCK & CO INC.  
 XX PI Anderson AS, Kuklin N, Jansen KU;  
 XX WPI; 2005-123069/13.

DR Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.  
 PS  
 PS Claim 7; SEQ ID NO 10; 84pp; English.

CC The present sequence is that of a *Staphylococcus aureus* protein ORF0657n  
 CC hybrid polypeptide denoted 0657nHybrid3. This is an example of claimed  
 CC hybrid polypeptide immunogens of the invention ADW88439-ADW88474 that  
 CC comprise a modified *S. aureus* ORF0657n sequence ADW88433-ADW88438  
 CC containing amino acid substitutions that increase sequence similarity to  
 CC ORF0190 ADW88432. The hybrid polypeptides contain one or more epitopes  
 CC for ORF0657n and ORF0190. They were designed by taking into account the  
 CC similarity and differences between native ORF0657n and ORF0190 protein  
 CC sequences. The invention also provides nucleic acids encoding these  
 CC hybrid polypeptides, and a method for evaluating the ability of an  
 CC immunogen to produce a protective immune response against *Staphylococcus*

CC infection using an animal (mouse or rat) model. The hybrid polypeptides  
 CC having therapeutic and diagnostic applications, such as being used to  
 CC provide protective immunity against *S. aureus* infection, being used to  
 CC generate antibodies to detect the presence of *S. aureus*, and being used  
 CC to generate therapeutic antibodies that target *S. aureus*.  
 XX  
 SQ Sequence 645 AA;

Query Match 13.7%; Score 89.5; DB 9; Length 645;  
 Best Local Similarity 29.4%; Pred. No. 3.4;  
 Matches 32; Conservative 19; Mismatches 43; Indels 15; Gaps 5;  
 QY 26 PEGKKDAGYVNLKSDTFIKPVFKIIEKKKEENKPTFDV-----SKKDNQPNVHNSQLN 80  
 DB 440 YEGQY---HVRIIDKDAFTKANTDKSNKKEQDSAKKREATPATPSKFTPSPEVKEBSOKQ 496  
 QY 81 ESHREKEDLQ---REHSQKSDTKDVT-ATVLDKNNISSKSTTNNPNK 124  
 DB 497 DSQKDDNKQLPSVKEKDASSESGDKTPTATKPTKGEVSSSTT--PTK 543

RESULT 38  
 ABP73992  
 ID ABP73992 standard; protein; 225 AA.  
 XX  
 AC ABP73992;  
 DT 30-JAN-2003 (first entry)  
 XX Candida albicans essential protein SEQ ID NO 7829.  
 DE Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
 KW signal transduction; DNA replication; cell division; growth;  
 KW proliferation; Candida albicans; fungicide; antifungal.  
 XX  
 OS Candida albicans.  
 XX WO200253728-A2.  
 XX 11-JUL-2002.  
 XX 26-DEC-2001; 2001WO-US049486.  
 XX 29-DEC-2000; 2000US-0259128P.  
 XX 20-FEB-2001; 2001US-00792024.  
 XX 22-AUG-2001; 2001US-0314050P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
 WPI; 2002-566694/60.  
 DR N-PSDB; ABZ32542.  
 DR  
 XX

CC Constructing strains for identifying gene products as effective targets  
 CC for therapeutic intervention, by inactivating in the strain one allele of  
 CC a gene and placing other allele of the gene under conditional expression.  
 XX  
 PS Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.

CC The invention relates to constructing (M1) a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified, comprising modifying  
 CC one allele by insertion or replacement by a cassette having an  
 CC expressible selectable marker and modifying other allele by  
 CC recombination, of a promoter replacement fragment with a heterologous  
 CC promoter, so that expression of the second allele is regulated by the  
 CC promoter. (M1) is useful for constructing a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified. The diploid fungal  
 CC cells having both alleles modified are useful for identifying a gene that  
 CC is essential to the survival or growth of a fungus, a gene that  
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
 CC that contributes to the resistance of a diploid fungus to an antifungal  
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus

CC and for identifying a therapeutic agent for treatment of a mammalian  
 CC disease. (M1) is useful for identifying a compound which modulates the  
 CC activity of a gene product, preferably enzymatic activity, carbon  
 CC compound catabolism, biosynthetic, transporter, transcriptional,  
 CC translational, signal transduction, DNA replication and cell division  
 CC activity. The method is useful for identifying a compound having the  
 CC ability to inhibit growth or proliferation of C. albicans cells and for  
 CC treating infection by C. albicans. The present sequence is that of an  
 CC essential Candida albicans protein used in the method of the invention.  
 CC Note: the sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office

XX  
 SQ Sequence 225 AA;

Query Match 13.5%; Score 88; DB 5; Length 225;  
 Best Local Similarity 28.0%; Pred. No. 1.2;  
 Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY 16 YQFDGWEISGPGKDGAVVNLSDTKTPVPKK----IREKKEENKPTFDVSKK-- 69  
 DB 80 YDDDDDEFGFSSNGAKELNLSQAIKWKQRDLIEBERKLNKKEEIIIRKAKS 139  
 QY 70 --DNPVNSQLNESHKEDLQREH--SQKSDSTKDVATVLDKNN 112  
 DB 140 TIDDFYENVNSKRDNHQKELLSEQKFIKRDDFLK--RGTLDNRVN 184

RESULT 39  
 AAB94584  
 ID AAB94584 standard; protein; 258 AA.

XX AAB94584;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:15383.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 15383; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893  
 CC represent human amino acid sequences, and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 258 AA;

Query Match 13.5%; Score 88; DB 4; Length 258;  
 Best Local Similarity 26.3%; Pred. No. 1.4;  
 Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;

QY 8 YKGELEKGYQFDGWEISGFGKGDAGY--VINLSKDTFIKVPFKIIEKKEENKPTFDV 65  
 DB 39 WEGEDEDEDVKNWDDDDDEKKEEANKVPEVISEK---KKAIEKIKERQKKRQBEI 95  
 QY 66 SKKXDNFQ----VNHSQLNESHKEDLQREHSSQKSDSTKDVATV--LDKNNISSK 116  
 DB 96 KKRLEPEEPKPVLTPEQLADKLRLKQLQESDLELAKETFGVNTVYGVGDAMPSSR 153

RESULT 40

ADP28113

ID ADP28113 standard; protein; 815 AA.

XX AC ADP28113;

XX 12-FEB-2004 (first entry)

XX Murine glycosyl-phosphatidyl-inositol-anchored protein homologue.

XX neuroprotective; nootropic; cerebroprotective; antiparkinsonian;

XX neurological; spinal cord injury; cranial; cerebral trauma; stroke;

XX Alzheimer's disease; anxiety; autism; Parkinson's; tardive dyskinesia;

XX paralysis; seizure; memory disorder; adiponectin; Ctg domain;

XX complement C1q; speract receptor; Wilm's tumour; synapsin; annexin;

XX leupin; serpin; NGRH; PRO; leucine-rich repeat; scavenger; neural IgCAM;

XX Ig; FN3; somatotropin; prolactin; somatostatin;

XX chorionic somatomammotropin hormone; NGAL; mucollipin; peroxidase; otx1;

XX SAPAP; murine; mouse.

XX Mus musculus.

XX WO2003048326-A2.

XX 12-JUN-2003.

XX 02-DEC-2002; 2002WO-US038526.

XX 03-DEC-2001; 2001US-00005499.

XX (HYSE-) HYSEQ INC.

XX Ghosh M, Tang YT, Wang JR, Wang Z, Zhao QH, Xu C, Mulero JJ;

XX Boyle BJ;

XX WPI; 2003-513756/48.

XX N-PSDB; ADF28753.

XX New polynucleotides and polypeptides, useful for useful for treating  
 XX neurological conditions, e.g. spinal cord injury, cranial or cerebral

PT trauma, stroke, Alzheimer's disease, anxiety, autism, Parkinson's disease, or paralysis.

PT disease, or paralysis.

XX  
PS Claim 9; SEQ ID NO 23; 396pp; English.

The invention relates to a novel isolated polynucleotide comprising any of the 87 sequences fully defined in the specification or its mature protein-coding portion. The polynucleotide of the invention demonstrates neuroprotective, neurotropic, cerebroprotective and antiparkinsonian activities whilst the polynucleotides, polypeptides and compounds may be useful for treating neurological conditions including spinal cord injury, cranial or cerebral trauma, stroke, Alzheimer's disease, anxiety, autism, Parkinson's disease, tardive dyskinesia, paralysis, seizures or memory disorders. The current sequence is that of the protein of the invention.

Sequence 815 AA:

**Query Match** 13.5% Score 88; DB 7; Length 815;

Query match	15.3%;	Score 88;	DB 7;	length 813;
Best Local Similarity	27.3%;	Pred. No. 6.7;		

Best local similarity 27.3%, Freq. NO: 0.7;  
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;

[illegible]

Db 196 GQEKQESFKSWBASGKHQEVSKPAVSLQQRKQDTSKLRSTLPREOKQOEISKSPSPSQ 255

QY 68 -KKDN PQVNHSQLNESHKED-----LQREBSQKSDSTKDVTATVLDKNISSKST 118

Db 256 WKQDTPKSKAGYVBEHKQETPKLPWPVLQKEQ-DPKQTPKSWTPSPMSEQNTTKSWT 314

Qy 119 T 119

Db 315 T 315

Search completed: April 24, 2006, 14:50:23

Job time : 81.8121 secs

**THIS PAGE BLANK (08PT0)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 81.3893 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTVSEEDFILPVYKG.....ATVLDKNISSKSTNNPNK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	696	100.0	2119	Q9AHT5_STRPN	Q9Aht5 streptococc
2	696	100.0	2140	Q97RY6_STRPN	Q97ry6 streptococc
3	693	99.6	2144	Q9S4M8_STRPN	Q9S4m8 streptococc
4	693	99.6	2144	Q8DQP7_STRR6	Q8dqp7 streptococc
5	125.5	18.0	300	Q4XU16_PLACH	Q4xui6 plasmodium
6	112	16.1	361	Q95P15_PLAFA	Q95pi5 plasmodium
7	110	15.8	346	Q9U0G0_PLARE	Q9u0g0 plasmodium
8	110	15.8	379	Q25705_PLAFA	Q25705 plasmodium
9	109	15.7	3008	Q8I436_PLAF7	Q8I436 plasmodium
10	108.5	15.6	600	Q77355_PLAF7	Q77355 plasmodium
11	108.5	15.6	1038	Q90784_CHICK	Q90784 gallus gall
12	108	15.5	384	Q250V0_ENTHI	Q250v0 entamoeba h
13	107	15.4	354	Q25995_PLAFA	Q25995 plasmodium
14	107	15.4	354	Q8I353_PLAF7	Q8I353 plasmodium
15	107	15.4	829	Q8I5F3_PLAF7	Q8I5f3 plasmodium
16	106	15.2	616	Q6BRW2_DREHA	Q6brw2 debaryomyce
17	103.5	14.9	379	Q8U6C4_PLAFA	Q8u6c4 plasmodium
18	103.5	14.9	380	Q26019_PLAFA	Q26019 plasmodium
19	103	14.8	296	Q250LX_ENTHI	Q250lx entamoeba h
20	102.5	14.7	379	Q25706_PLAFA	Q25706 plasmodium
21	102	14.7	540	Q54MT2_DICDI	Q54mt2 dictyosteli
22	101.5	14.6	382	Q9V7J0_DROME	Q9v7j0 drosophila
23	101.5	14.6	556	Q9V7I9_DROME	Q9v7i9 drosophila
24	101.5	14.6	785	Q9GQ82_DROME	Q9gq82 drosophila
25	101.5	14.6	954	Q2GHNO_BACHK	Q2ghno bacillus th
26	101	14.5	662	Q4YMU4_PLABE	Q4ymu4 plasmodium
27	100	14.4	329	Q9NFV9_PLACH	Q9nf9v plasmodium
28	100	14.4	1011	Q4Y2I3_PLACH	Q4y2i3 plasmodium
29	100	14.4	1130	Q8IJZ4_PLAF7	Q8Ijz4 plasmodium
30	99.5	14.3	374	Q9V9M0_PLAKN	Q9v9m0 plasmodium
31	99	14.2	211	P91488_CABEL	P91488 caenorhabdi

32 99 14.2 437 2 Q54K26\_DICDI Q54k26 dictyosteli  
33 99 14.2 467 2 Q59PE2\_CANAL Q59pe2 candida alb  
34 99 14.2 467 2 Q59PL2\_CANAL Q59pl2 candida alb  
35 99 14.2 1028 2 Q4Z4Q1\_PLABE Q4z4q1 plasmodium  
36 99 14.2 1859 2 Q8IC27\_PLAF7 Q8ic27 plasmodium  
37 99 14.2 2563 2 Q8I3A0\_PLAF7 Q8i3a0 plasmodium  
38 98.5 14.2 393 2 Q7RKU2\_PLAYO Q7rku2 plasmodium  
39 98.5 14.2 674 2 Q7RLR7\_PLAYO Q7rlr7 plasmodium  
40 98.5 14.2 827 2 Q5SG46\_DICDI Q5sg46 dictyosteli  
41 98 14.1 157 2 Q9VQV0\_DROME Q9vgv0 drosophila  
42 98 14.1 951 2 Q8I229\_PLAF7 Q8i229 plasmodium  
43 97.5 14.0 449 2 Q8IHW3\_PLAF7 Q8ihw3 plasmodium  
44 97.5 14.0 556 2 Q95S93\_DROME Q95s93 drosophila  
45 97.5 14.0 1345 1 YHO0\_YEAST P38800 saccharomyc

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RX DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Magure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RA "Use of a whole genome approach to identify vaccine molecules  
RT affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SST.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 5179B7F6B960A6A CRC64;

Query Match 100.0%; Score 696; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 4.6e-43;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 60
Db 1952 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 2011
Qy 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
Db 2012 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2071
Qy 121 NNISKSSTNNPNK 134
Db 2072 NNISKSSTNNPNK 2085

RESULT 2
Q37RY6 STRPN
ID Q97RY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AB007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSP; P00782; 2SST.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRfams; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRfams; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; F744AD8E2938B334 CRC64;

Query Match 100.0%; Score 696; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 4.7e-43;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 60
Db 1973 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 2032
Qy 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
Db 2033 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092
Qy 121 NNISKSSTNNPNK 134
Db 2093 NNISKSSTNNPNK 2106

RESULT 3
Q9S4M8 STRPN
ID Q9S4M8_STRPN PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor PrtA.
GN Name=PrtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3.B;
RX MEDLINE=21585565; PubMed=11728722;
RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
RA Zysk G.;
RT "The cell wall-associated serine proteinase PrtA: a highly conserved
RT virulence factor of Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 205:199-204(2001).
DR EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
DR HSP; P00782; 2SST.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRfams; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```

DR PROSITE, PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Cell wall; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 2144 cell wall-associated serine proteinase
FT PRTA.
SQ SEQUENCE 2144 AA; 240725 MW; 20525114707411331 CRC64;

Query Match 99.6%; Score 693; DB 2; Length 2144;
Best Local Similarity 99.3%; Pred. No. 7.8e-43;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFIKPVFKK 60
DB 1977 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFIKPVFKK 2036

QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDIQREHSHQKSDTKDVTATVLDK 120
DB 2037 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDIQREHSHQKSDTKDVTATVLDK 2096

QY 121 NNISKSTTNNPNK 134
DB 2097 NNISKSTTNNPNK 2110

RESULT 4
QBDQP7_STRR6
ID Q8DQP7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocusNames=spr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Mateuuehima P.,
RA McLaren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peary R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AK93365.1; -; Genomic_DNA.
DR F1; A97942; A97942.
DR HSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0036020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042803; F:protein self binding; IEA.
DR GO; GO:004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR00209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.

```

```

DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 693; DB 2; Length 2144;
Best Local Similarity 99.3%; Pred. No. 7.8e-43;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFIKPVFKK 60
DB 1977 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFIKPVFKK 2036

QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDIQREHSHQKSDTKDVTATVLDK 120
DB 2037 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDIQREHSHQKSDTKDVTATVLDK 2096

QY 121 NNISKSTTNNPNK 134
DB 2097 NNISKSTTNNPNK 2110

RESULT 5
Q4XUI6_PLACH
ID Q4XUI6_PLACH PRELIMINARY; PRT; 300 AA.
AC Q4XUI6;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PC000286.03.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAATJ01003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD0000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 4.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 300 AA; 34469 MW; 8ED4E512AFB1945 CRC64;

Query Match 18.0%; Score 125.5; DB 2; Length 300;
Best Local Similarity 28.7%; Pred. No. 0.18;
Matches 33; Conservative 26; Mismatches 31; Indels 25; Gaps 6;

QY 15 LPVYKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFIK-----PVFKIEBKKE 66

```



```
Db 164 LAVYRGHVPAYKI-AWSI-----DNNYIVSCSDSTLKLRLINHLVPLLKKEENAE 215
Qy 67 ----BENKTFDVSCKKQNPVNHSQLNESHKEDLQREHSQKSDSTKQVATVL 118
Db 216 QTKDQK-----NEQKENPQ-NNDQPNDEANSEKKCKKCKEKNKDKTKNKIKTKLL 264

RESULT 6
Q5P15 PLAPA
ID Q5P15 PLAPA PRELIMINARY; PRT; 361 AA.
AC Q5P15;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-PVO;
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;
RA Hisaeda H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,
RA Stowers A.W.;
RT "Merozoite surface protein 3 and protection against malaria in Aotus
RT nancymai monkeys";
RL J. Infect. Dis. 185:657-664 (2002).
DR EMBL; AY04180; AAK94780.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT CHAIN <1> >361 merozoite surface protein 3.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 16.1%; Score 112; DB 2; Length 361;
Best Local Similarity 23.9%; Pred. No. 2.2;
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;

Qy 18 YKGELEKGYQ-----PD---GWEISGF--EGKQDAG-----YVI 46
Db 165 YAGVKEDYERAKNAVQKANQAVLKAKKASSDYILGWFGGVPFHKKEENMLSHLYS 224
Qy 47 NLSKDTFTFKPVFKKEEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQK 106
Db 225 SKOKENISKENDVDLDE-KEEAETETEEELKEKNEETETSEISEDEEEEEEKEE 283
Qy 107 SDSTKQVATVLQKN-----NISKSTTN 131
Db 284 NDKKKEQKEQSNENNNDQKQMEANLISKQNNN 318

RESULT 7
Q5U0G0 PLARE
ID Q5U0G0 PLARE PRELIMINARY; PRT; 346 AA.
AC Q5U0G0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188 (2000).
```

```
DR EMBL; AJ252286; CAB65754.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON_TER 1 1
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAFAP010 CRC64;

Query Match 15.8%; Score 110; DB 2; Length 346;
Best Local Similarity 26.2%; Pred. No. 2.9;
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;

Qy 17 VYKGELEKGYQFD-GWEISGF--EGKQDAG-----YVINLSKDTFTFKPVFKKEEKEE 68
Db 184 VLKAKKASSYNYILGWFGGVPFHKKEENMLSHLYSVSKKENISKENDVDLDE-KEE 242
Qy 69 NKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQKSDSTKQVATV-----LDKN 121
Db 243 AEETGEQELKEKNEETETSEINEDEFOEBEEEEEKEEENNDKKEQAKESQNDQKEDME 302
Qy 122 NISKSTTN 131
Db 303 NLISKQNNN 312

RESULT 8
Q25705 PLAPA
ID Q25705 PLAPA PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31 (1997).
DR EMBL; U08851; AAC47831.1; -; Unassigned_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 15.8%; Score 110; DB 2; Length 379;
Best Local Similarity 23.4%; Pred. No. 3.2;
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

Qy 18 YKGELEKGYQ-----PD---GWEISGF--EGKQDAG-----YVI 46
Db 185 YAGVKEDYERAKNAVQKANQAVLKAKKASSDYILGWFGGVPFHKKEENMLSHLYS 244
Qy 47 NLSKDTFTFKPVFKKEEENKPTFDVSKKQNPVNHSQLN-----ESH 93
Db 245 SKOKENISKENDVDLDE-KEEAETETEEELKEKNEETETSEISEDEEEEEEKEEENE 303
Qy 94 RKEDLQREHSQKSDSTKQVATVLQKNISKSTTN 131
Db 304 KKEQKEQESQNNNDQKQMEANLISKQNNN 336

RESULT 9
Q81436 PLAF7
ID Q81436 PLAF7 PRELIMINARY; PRT; 3008 AA.
AC Q81436;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PFE0325w.
```

GN Name=PF0325w;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3D7;  
EX MEDLINE=2255708; PubMed=12368867; DOI=10.1038/nature01095;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
RA Sulston J.E., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3D7;  
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,  
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL929351; CAD51431.1; -: Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBEE15C599B4 CRC64;  
  
Query Match 15.7%; Score 109; DB 2; Length 3008;  
Best Local Similarity 32.4%; Pred. No. 35;  
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;  
  
Qy 44 YVINSK----DTFTKPKKIEKKKEENKPTFDVSKKKNQPNVHQSQ---LNESHKKE 96  
Db 2310 YDIELSKLEKGCASIGPFTD--EENKKEENK--EVNKKKEENKKEENKKEENKKE 2366  
  
Qy 97 DLQREH----SQKSDSTKVATVLDKRNISK-----STNNPNK 134  
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKE 2414  
  
RESULT 10  
O77355 PLAF7 PRELIMINARY; PRT; 600 AA.  
AC O77355;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DE Hypothetical protein MAL3P4.20.  
OS Name=MAL3P4.20; Synonyms=PF0465c;  
GN Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
RA Horrocks P., Jagels K., Jaseal B., Kyes S., Mclean J., Moule S.,  
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum.";  
RL Nature 400:532-538 (1999).

RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=2255708; PubMed=12368867; DOI=10.1038/nature01095;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
RA Sulston J.E., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3D7;  
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,  
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL929351; CAD51431.1; -: Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;  
  
Query Match 15.6%; Score 108.5; DB 2; Length 600;  
Best Local Similarity 29.3%; Pred. No. 6.8;  
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;  
  
Qy 7 IVSEEDILPVY-----KGELEKGYQDQWEIFSGEKK-----DAGYVINISKDTPIKPV 57  
Db 60 ILGFEDILYCYISQLKQSKKK---DGBEDKYLNAKCKLKNLTGFTGNKKSDFIBEL 116  
  
Qy 58 FKKI--BEKKEE-----ENKPTFDVSK--KKNQPNVHQSQNE-----SHRK 95  
Db 117 LELLNEKKEEHTADTLNENK--TDYKKNVNNENYNNYNNYNNYNNYNNYNNYNNYNNYNN 175  
  
Qy 96 E-----DLQREH-----SQKSDSTK-----DVTATVLDKRNISKSTTN 130  
Db 176 EHNINNVNKKKEYTDTIQDKRKHESLSQKSDSYKKRPKNKKTSTIER--SLSNKRYDE 234  
  
Qy 131 NPNK 134  
Db 235 KTNK 238  
  
RESULT 11  
Q90784 CHICK PRELIMINARY; PRT; 1038 AA.  
AC Q90784;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Claustrian.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=94157526; PubMed=7906711;  
RA Burg M.A., Cole G.J.;  
RT "Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally related to MAP1B.";  
RL J. Neurobiol. 25:1-22(1994).



RESULT 16  
 Q6BWR2\_DBBHA  
 ID Q6BWR2\_DBBHA PRELIMINARY; PRT; 616 AA.  
 AC Q6BWR2;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Similar to CA4458|IPF8464 Candida albicans IPF8464 unknown function.  
 DE  
 GN OrderedLocusNames=DEHA0D14674g;  
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
 NCBI\_TaxID=4959;  
 RX NCBI\_TaxID=4959;  
 RN [1]  
 NU NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=ATCC 36239 / CBS 767;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
 RA Goffard N., Frangoul L., Algile M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boisrame A., Boyer J., Catcolico L., Confaniolieri F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekia F., Wesolowski-Leuvel M., Westhof E., Wirth B.,  
 RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.-L.;  
 RL "Genome evolution in yeasts."  
 RT Nature 430:35-44(2004).  
 DR EMBL: CR382136; CAG87226.1; -: Genomic DNA.  
 DR GO: GO:0016303; F:kinase activity; IEA.  
 DR InterPro: IPR000749; ATP\_gua\_Ptrans.  
 KW Complete proteome.  
 SQ SEQUENCE 616 AA; 72143 MW; 884009B2B8B6C3CF CRC64;  
  
 Query Match 15.2%; Score 106; DB 2; Length 616;  
 Best Local Similarity 30.7%; Pred. No. 11;  
 Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9  
  
 QY 16 PVYKGELEKGYFDGWEISGPEGKADAGYVNLTKDT-FIKPV----FKIIEBKGEENK 70  
 DB 87 PVLKGRRAKFKYITLDTI-----KQIN-DINFSDESEHEKPIETSKKEKTKYTK 140  
 QY 71 PTFDVSKKK-----DNQV-VNHSQALNE---SHRKDLQREHSQK-----SDSTKQVT 114  
 DB 141 PLDIDIGKLRRIADNPDIENHSSESEIKQRKEKQRQKQDKRKLKANKQCESNNDST 200  
 QY 115 ATVLDK-NNISSKSTNNPN 133  
 DB 201 TQSPBLKNNINEKITSNEPS 220  
  
 RESULT 17  
 Q9U6C4\_PLAFA  
 ID Q9U6C4\_PLAFA PRELIMINARY; PRT; 379 AA.  
 AC Q9U6C4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Polymorphic antigen.  
 GN Name=WSP-3;  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=5833;  
 RX [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=PCCC/HN;  
 RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;

Db	312 EQSNENNDQKKOMEA	Q50LX8_ENTHI	Q50LX8_ENTHI PRELIMINARY	PRT	296 AA
Q50LX8_ENTHI	Q50LX8_ENTHI PRELIMINARY	PRT	296 AA		
AC	Q50LX8				
DT	13-SEP-2005 (T-EMBLrel. 31, Created)				
DT	13-SEP-2005 (T-EMBLrel. 31, Last sequence update)				
DT	13-SEP-2005 (T-EMBLrel. 31, Last annotation update)				
DE	Hypothetical protein.				
GN	ORFNames=657.t00001;				
OS	Entamoeba histolytica HM-1:IMSS.				
OC	Eukaryota; Entamoebidae; Entamoeba.				
OX	NCBI_TaxID=294381;				
RP	[1]				
RN	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=HM-1:IMSS;				
RC	PubMed=15729342; DOI=10.1038/nature03291;				
RA	Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J., Amedeo P., Roncaglia P., Beriman M., Hirt R.P., Mann B.J., Nozaki T., Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M., Hofer M., Bruchhaus I., Willhoft U., Bhattacharya A., Chillingworth T., Churcher C., Hanze Z., Harris D., Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S., Quail M.A., Rabinowitz E., Norbartzak H., Price C., Wang Z., Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A., Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C., El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B., Fraser C.M., Hall N.;				
RA	"The genome of the protist parasite Entamoeba histolytica.";				
RL	Nature 433:865-868(2005).				
CC	-1- CAUTION: The sequence shown here is derived from an preliminary data.				
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is				
DR	EMBL; AAFB01001439; EAL42595.1; -; Genomic_DNA.				
KW	Hypothetical protein.				
SQ	SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;				
Query Match	14.8%; Score 103; DB 2; Length 296;				
Best Local Similarity	28.4%; Pred. No. 8.2;				
Matches	40; Conservative 23; Mismatches 50; Indels 28; Gaps				
Qy	1 KEMSTTVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGADYVYINLSKDTFKIPVFKK 60				
Db	135 KEQKTEKSE-----GDSEKK-----DIPTEGKKNK-----DITKDKNDKEKD 177				
Qy	61 IEKKE---ENKPTDFVSKKK-DNPQVNH-----QLNFSHRKEDLQREHSQKSDS 109				
Db	178 TNEEGSSGKGQKTEESKCTENPQNNSSNKKEEKKKEEKKKEEKKKEEKKKEE 237				
Qy	110 TKDVTATVLDKNNISSKSTTN 130				
Db	238 EQNQKPEINEKNEKNTTN 258				
RESULT 20					
Q25706_PLAFA	Q25706_PLAFA PRELIMINARY	PRT	379 AA		
AC	Q25706				
DT	01-NOV-1996 (T-EMBLrel. 01, Created)				
DT	01-NOV-1996 (T-EMBLrel. 01, Last sequence update)				
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)				
DE	Polymorphic antigen.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RP	[1]				
RN	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=FC27;				
RC	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/S0166-6851(97)00130-8;				
RA	McColl D.J., Silva A., Foley M., Kun J.F., Ravaloro J.M., Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;				
RT	"Molecular variation in a novel polymorphic antigen associated with Plasmodium falciparum merozoites.";				
RL	Mol. Biochem. Parasitol. 68:53-67(1994).				
CC	[2]				
CC	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=FC27;				
RC	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/S0166-6851(97)00130-8;				
RA	McColl D.J., Silva A., Foley M., Kun J.F., Ravaloro J.M., Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;				
RT	"Molecular variation in a novel polymorphic antigen associated with Plasmodium falciparum merozoites.";				
RL	Mol. Biochem. Parasitol. 68:53-67(1994).				
CC	[2]				
CC	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=FC27;				
RC	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/S0166-6851(97)00130-8;				
RA	McColl D.J., Silva A., Foley M., Kun J.F., Ravaloro J.M., Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;				
RT	"Molecular variation in a novel polymorphic antigen associated with Plasmodium falciparum merozoites.";				
RL	Mol. Biochem. Parasitol. 68:53-67(1994).				
CC	[2]				
CC	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=FC27;				
RC	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/S0166-6851(97)00130-8;				
RA	McColl D.J., Silva A., Foley M., Kun J.F., Ravaloro J.M., Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;				
RT	"Molecular variation in a novel polymorphic antigen associated with Plasmodium falciparum merozoites.";				
RL	Mol. Biochem. Parasitol. 68:53-67(1994).				
CC	[2]				
CC	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=FC27;				
RC	MEDLINE=95198774; PubMed=7891748; DOI=10.1016/S0166-6				





```

RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield J.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirkas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved
isoform of Asph missing the catalytic domain share exons with
junctin.";
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF2003808; AAF58063.2; -; Genomic_DNA.
DR EMBL; AF289494; AAG40807.1; -; mRNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 382 AA; 43287 MW; 60E5C03AEBFC6E8B CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 382;
Best Local Similarity 24.5%; Pred. No. 14;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDPLPVYGELEKGYQPGW-----ETSGEGKKDAGVY-----NLKDTFTK 55
Db 78 EDLDTLSERSRK--VFDGWDEHDEHDGHVQPSGEALDDHDEHDDHDDHDEDEE 135
QY 56 PVFKIEEKEENKPT-----FDVSKKDDPQVNHSQLNSHREKDLQREHSQKSDS 109
Db 136 PLTEELLEELEEEETDEPADEYEDEDEENNA--GENITAEADAEDEEDND 193

QY 110 TKDVTATVDKNNISKST 128
Db 194 EGTVEATVEATTEATT 212

RESULT 23
Q9V719 DROME
ID Q9V719 DROME PRELIMINARY; PRT; 556 AA.
AC Q9V719;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
```

```

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DB CG8421-PD, Isoform D (CG8421-pe, isoform e).
GN Name=Asph; ORFNames=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hoadic K., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
```





```
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Ceiniker S., Carlson J., Wan K., Pfeiffer B., Fries E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289493; AAG40806.1; -; mRNA.
DR EMBL; AE003808; AAM70947.1; -; Genomic DNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IEA.
DR GO; GO:0056021; C:integral to membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; IEA.
DR GO; GO:0018193; F:peptidyl-amino acid modification; IEA.
DR InterPro; IPR007803; Asp Arg Hydrex.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001440; TPR-like helical.
DR InterPro; IPR011990; TPR-like helical.
DR Pfam; PF05118; Asp Arg Hydrex; 1.
DR PROSITE; PS50293; TPR REGION; 1.
DR PROSITE; PS00142; ZINC PROTEASE; UNKNOWN 1.
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFC6836F7F1 CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 785;
Best Local Similarity 24.5%; Pred.No. 30;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDFILPVYKGELEKGVQFDGW-----EISGEGKKDAGYVI-----NLSKDTIK 55
DB 78 ELDLTFLESRSFK-VFDGWDEHRDHDGHDVQSPGEALDDHDEHDDHDDHDEDEE 135
QY 56 PVFKIEEKKKEENKPT-----FDVSKKKNPQVNHSQLNESHKRDLPQREHSQKSDS 109
DB 136 PLTEELLEELEERPTDEPADEAEVEEDEDENNA--GENITAEADAEEREDND 193
QY 110 TKDVTATVLDKNISKST 128
DB 194 EGTVEATVEATTEATTEAT 212

RESULT 25
QHNR0 BACHK
ID Q6HNR0 BACHK PRELIMINARY; PRT; 954 AA.
AC Q6HNR0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Possible internalin protein.
GN OrderedLocustNames=BT9727_0463;
OS Bacillus thuringiensis (subsp. konkukian).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
ON NCBI_TaxID=180856;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=97-27;
RA Bretin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus thuringiensis 97-27."
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017355; AAT63966.1; -; Genomic DNA.
DR GO; GO:0009986; C:cell surface; IEA.
```

```
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003885; LRR_cyst.
DR InterPro; IPR007092; LRR_SDS22.
DR InterPro; IPR006635; NEA_transpt.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00560; LRR_1; 8.
DR Pfam; PF05031; NEAT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00365; LRR_SD22; 8.
DR SMART; SM00725; NEAT; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50978; NEAT; 1.
KW Complete proteome.
SQ SEQUENCE 954 AA; 108589 MW; 4F4CF8B44C9B355F CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 954;
Best Local Similarity 25.7%; Pred.No. 37;
Matches 43; Conservative 32; Mismatches 43; Indels 49; Gaps 9;

QY 11 EDFILPVY--KGEL-----EKGYP--DGWEIS--GFEG--- 38
DB 727 ERIQVPYDLEGEIENIKUTSDGTNNGVKWTGPEKVKYKFDLSDSEISFNQTVI 786
QY 39 -----KKDAGYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLN 90
DB 787 QNIVEKEEKEPTKEVEESKEEKEPT-KEVEESKEEKEPTKEVEESKEEVEPTKEVE 845
QY 91 ESHKEDLPQREHSQKSDSTKYDVTATVLDKN-----NLSKSTTNPN 133
DB 846 ES--KEEV--KEPTKEVEESKEEVAJEIIEKKEEINQSAFVQEQNVN 888

RESULT 26
Q4YMU4_PLABE
ID Q4YMU4_PLABE PRELIMINARY; PRT; 662 AA.
AC Q4YMU4;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE RNA binding protein, putative.
GN ORFNames=PB001104.03.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5821;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAI01003467; CAI00666.1; -; Genomic_DNA.
DR InterPro; IPR012972; NLE.
DR InterPro; IPR001680; WD40.
DR Pfam; PF08154; NLE; 1.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 662 AA; 75536 MW; E7C8543AF5E59124 CRC64;
```

OX	NCBI_TaxID=5825;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RA	Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA	Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA	James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA	Quail M.A., Ormond D., Doggett J., Trueman H.S., Mendoza J.,
RA	Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA	Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.,
RT	"A comprehensive survey of the Plasmodium life cycle by genomic,
RT	transcriptomic, and proteomic analyses.";
RL	Science 307:82-86(2005).
CC	-1- CAUTION: The sequence shown here is derived from an
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
CC	EWBL: CAJAJ01001570; CAH76627.1; -; Genomic_DNA.
DR	Hypothetical protein.
KW	NON_TER 1011 1011
FT	SEQUENCE 1011 AA; 119345 MW; 0D8893E7EB59DADS CRC64;
SQ	
	Query Match 14.4%; Score 100; DB 2; Length 1011;
	Best Local Similarity 28.7%; Pred. No. 51;
	Matches 35; Conservative 26; Mismatches 43; Indels 18; Gaps 6;
QY	14 ILPVYGELEKGYQFGWISGPEG--KKDAGYV--NLKOTFIKVPFKIEKKEEN 69
DB	368 ITPI-KNEIDIVENYDKDISIFONICQKENGNIIGENSHKET-NKYIHNMKIKNSISK 425
QY	70 KPTFDYKKKONPQVNHSQLNSHRKEDLQREHSQKSDSTKDVATVLDKNISSKTT 129
DB	426 KNEMLKKKNDKQKQSHFNQ-----KTKNSD--KNTNATYINETNLNSYNT 473
QY	130 NN 131
DB	474 SD 475
	;;
RESULT 29	
Q81J24_PLAF7	
ID	Q81J24_PLAF7 PRELIMINARY; PRT; 1130 AA.
AC	Q81J24;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	Hypothetical protein.
OS	ChFNames=PF10_0046;
OS	Plasmodium falciparum (isolate 3D7).
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX	NCBI_TaxID=36329;
RP	[1]
RN	NUCLEOTIDE SEQUENCE.
RL	MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA	Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA	Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA	Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA	Partea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA	Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA	McRadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA	Fraser C.M., Barrell B.G.;
RT	"genome sequence of the human malaria parasite Plasmodium
RT	falciparum";
RL	Nature 419:498-511(2002).
CC	-1- SIMILARITY: Contains 1 RING-type zinc finger.
DR	EWBL: AB014829; AAN3244.1; -; Genomic DNA.
DR	GO: GO:0000151; C:ubiquitin ligase complex; IEA.
DR	GO: GO:0046872; F:metal ion binding; IEA.
DR	GO: GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR	GO: GO:0008270; F:zinc ion binding; IEA.
DR	GO: GO:0016567; P:protein ubiquitination; IEA.
DR	InterPro: IPR006209; EGF like.
DR	InterPro: IPR001841; ZnF RING.

```
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Hypothetical protein; Metal-binding; Ubl conjugation pathway; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1130 AA; 131698 MW; FE6AAFE08C4CCDB6 CRC64;

Query Match 14.4%; Score 100; DB 2; Length 1130;
Best Local Similarity 26.1%; Pred. No. 57;
Matches 30; Conservative 18; Mismatches 53; Indels 14; Gaps 3;

QY 34 SGFEKKDAGVY--INLSKDTPIKVPFKIEKKEEENKPTFDVSKKKNPQVNHSQLNE 91
Db 568 SFLSGNSDYSRIDRYLSPDNYPVSNNRNINDENVVEIVQZETSDYNESHDEYNE 627
QY 92 SHRKEDLQREHSQKS-----DSTKDVATVLDKN-NISSKSTTTNNPNK 134
Db 628 NNEKEINREGNEBEKSKRKKTNDYDNDKDVYNNASGNSNHSNNNNNNNK 682

RESULT 30
QSV9M0_PLAKN PRELIMINARY; PRT; 374 AA.
AC QSV9M0;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Merozoite surface protein 5.
GN Name=MSP5;
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RC STRAIN=H;
RA Black C.G., Wang L., Topolska A.E., Finkelstein D.I., Horne M.K.,
RA Thomas A.W., Mohandas N., Coppel R.L.;
RT "Merozoite surface proteins 4 and 5 of Plasmodium knowlesi have
RT differing cellular localisation and association with lipid rafts.";
RL Mol. Biochem. Parasitol. 138:153-158(2004).
DR EMBL; AY573058; AAT77929.1; -; Genomic_DNA.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 374 AA; 41642 MW; COD687C6F23FE989 CRC64;

Query Match 14.3%; Score 99.5; DB 2; Length 374;
Best Local Similarity 26.6%; Pred. No. 19;
Matches 37; Conservative 25; Mismatches 56; Indels 21; Gaps 6;

QY 10 BEDFILPVYKGL--EKG-----YFDGWEISGFGKKDAGVYINL-----SKDTPIK 55
Db 31 EEKHILQIYTNRLLEEGKNDNVQWESPSISGTGKEIKQIMISHLQLQSGKHDOVSPLS 90
QY 56 PVFKIEEKEENKPTFDVSKKKNPQVNHSQLNEHSHKEDLQREHSQKSDSKDTVA 115
Db 91 ANDSNLKGNEBASAESGNGKSDENNVKQSD--ENAKKSD---EENKANSNTKDAES 146
QY 116 TVLDDKNNISSK---TTNN 131
Db 147 AGEENPVSQENQMTLNN 165

RESULT 31
P91488_CABEL PRELIMINARY; PRT; 211 AA.
AC P91488;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein T23B3.5.
GN ORFNames=T23B3.5;

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RC MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; U88309; AAB42334.1; -; Genomic_DNA.
DR PIR; T25911; T25911.
DR Ensembl; T23B3.5; Caenorhabditis elegans.
DR WormBase; WBGene00020713; T23B3.5.
DR WormPep; T23B3.5; CE14016.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 211 AA; 23956 MW; 11B16164A87B5928 CRC64;

Query Match 14.2%; Score 99; DB 2; Length 211;
Best Local Similarity 29.4%; Pred. No. 11;
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 37 EGKKDAGVINLSKDTPIKVPFKIEKKEEENKPTFDVSKKKNPQVNH----SQLNES 92
Db 69 EGKKDGEKSEKKGDKGKKEEKKDKKDEKDKKDEKDKKDEKDKKDEKDKDE 128
QY 93 HRKEDLQREHSQKSDSKDTVAIVLDKNNISSTTTNNPNK 134
Db 129 EKDDKDEKDEKDEKDEKDEKDEKSKSKSKSKSKSKSK 170

RESULT 32
Q54K26_DICDI PRELIMINARY; PRT; 437 AA.
AC Q54K26;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=DDB0219257;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Turggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Fabrother J., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hausser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulsegged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaubky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF10100156; EAL63646.1; -; Genomic_DNA.
KW Hypothetical protein.
```



```

NCBI_TaxID=5821;
[1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford L., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Jase C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RA "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CA0100607; CAH94732.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1028 AA; 120493 MW; 866A9FFPCC427612 CRC64;
Query Match 14.2%; Score 99; DB 2; Length 1028;
Best Local Similarity 25.0%; Pred. No. 62;
Matches 41; Conservative 22; Mismatches 53; Indels 48; Gaps 6;
QY 10 BEDFILPVYKGLKGYQFDGWEISGFGKKDAGYVINSKDTFTKPVFKK----- 60
Db 566 DDDYYDYDESYEYKGEK-----KKFVFEKGN-----NLKNNKKEKLNKGKNNKINE 616
QY 61 ---IEBKGEENKPTFDVSKKKNPNQVNSHLSKSHKEDLQRE-----EHSOKS 107
Db 617 IVTKGESENKKEAYDETNNKIKKSKNDLVEPKKIKBEIKNEYIIDENKGRKQTEKGEKE 676
QY 108 DS-----TKDVTATVLDKNNI-----SSKSTTNNPK 134
Db 677 NKNNSIKKYKLTFOFEEHFLKEDNIEIGKKNESSTKKNNSNK 720
RESULT 36
Q8IC27_PLAF7
ID Q8IC27_PLAF7 PRELIMINARY; PRT; 1859 AA.
AC Q8IC27;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein PF07_0016.
GN Name=PF07_0016;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RL EMBL; AL844506; CAD50814.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1859 AA; 219375 MW; 2266544164ED360C CRC64;
Query Match 14.2%; Score 99; DB 2; Length 1859;
Best Local Similarity 24.7%; Pred. No. 1.2e+02;
Matches 36; Conservative 24; Mismatches 60; Indels 26; Gaps 5;
QY 6 TIVSEDFILPVYKGLKGYQFDGWEISGP-----EGKKDAGYVINSKDT- 52
Db 216 TDYSEDEIYAKYIQDKSSDNYSQGYDKSKLINTSNINMLNVTNKNQVNH--SMSNTI 273
QY 53 -----PIKPVFKIKIEKKEENKPTFDVSKKKNPNQVNSHLSKSHKEDLQREHSOKS 107
Db 274 QQDLSFIHSINKYEKKKEKKNK-NYDRNKKSSNTNDKSYNTQNDPRKNQKQFVDNN 332
QY 108 DSTKDVATVLDKNNISSKSTTNNPN 133
Db 333 NKRND-----HNKNNELEQVYNNPN 353
```

```

RESULT 37
Q8I3A0_PLAF7
ID Q8I3A0_PLAF7 PRELIMINARY; PRT; 2563 AA.
AC Q8I3A0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE E1-E2 ATPase/hydrolase, putative.
GN Name=PF10240c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbintowitsch E.,
RA Rajandream M.A., Rutter S., Smith R., Squares R., Squares M., Simmonds M.,
RA Seeger K., Sharp S., Taylor K., Unwin L., Whitehead S., Stevens K.,
RA Taylor K., Thivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL2929355; CAD51734.1; -; Genomic_DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0016820; F:hydrolase activity, acting on acid anhydrid. . .; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR GO; GO:0030001; P:metal ion transport; IEA.
DR InterPro; IPR001757; ATPase E1-E2.
DR InterPro; IPR005834; Dehal like hydro.
DR InterPro; IPR008250; E1-E2_ATPase_reg.
DR InterPro; IPR008121; HeavyWe_transpt.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolase; 2.
DR PRINTS; PR00119; CATATPASE.
DR TIGRFAMs; TIGR01494; ATPase P-type; 1.
DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
DR PROSITE; PS50846; FMA_2; 1.
KW Hydrolase.
SQ SEQUENCE 2563 AA; 298806 MW; 3F9613243D26F9F1 CRC64;
Query Match 14.2%; Score 99; DB 2; Length 2563;
Best Local Similarity 29.4%; Pred. No. 1.6e+02;
Matches 35; Conservative 18; Mismatches 60; Indels 6; Gaps 4;
QY 20 GELEKGYQFDG-WEISGFGKKDAGYVINSKDTFTKPVFKKIEKKEENKPTFDVSKK 78
Db 856 GEEKKG-NIDGIYILKQNHKKOMIKGEENKDNFNSKEEKSDNSENNEIDKNYNLKR 914
QY 79 KD----NQVNSHLSKSHKEDLQRE-HSOKSDSTKDVATVLDKNNISSKSTTNNPN 133
Db 915 KEKHNSNDNDNDSVLNKNVKEEITHESSESSNEQSDSYLKKIEKKNKNSVDNNEN 973
RESULT 38
Q7RKU2_PLAYO
ID Q7RKU2_PLAYO PRELIMINARY; PRT; 393 AA.
AC Q7RKU2;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
```



DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE Drosophila melanogaster G12259 gene product.  
 OS Name=PY02808;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=17XNL;  
 RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perteau M.,  
 RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC EMBL: AABL01000781; EAA22302.1; -; Genomic\_DNA.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR InterPro: IPR007005; XAP5.  
 DR PANTHER: PTHR12722; XAP5; 2.  
 DR Pfam: PF04921; XAP5; 1.  
 SQ SEQUENCE 393 AA; 46652 MW; 55B30519B8FA97D2 CRC64;  
  
 Query Match 14.2%; Score 98.5; DB 2; Length 393;  
 Best Local Similarity 32.7%; Pred. No. 24;  
 Matches 35; Conservative 14; Mismatches 47; Indels 11; Gaps 5;  
  
 QY 36 FEGKDGAGVNLNLSKDT-----PIKPVFKIEKKEENKPTPDVSKKDNPOVNHQS 88  
 DB 90 FSKEDTTY-NESSKDTYKHKHTNFKLSPFSDDEEDEDKNDENKSETPK-NKSD 147  
 QY 89 LNESHRKEDLQREHSQKSDSTKVTVTLDRNNISSKS-TTNNPNK 134  
 DB 148 EN-SLEKEQNEKEAEKSNSETEQINKNTYDKNLQNGKSVNTENK 193  
  
 RESULT 39  
 Q7RL87 PLAYO PRELIMINARY; PRT; 674 AA.  
 AC Q7RL87;  
 DT 01-MAR-2004 (TReMBLrel. 26, Created)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)  
 DE Nochless-related.  
 GN Name=PY02598;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=17XNL;  
 RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perteau M.,  
 RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;

RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC EMBL: AABL01000712; EAA22065.1; -; Genomic\_DNA.  
 DR HSSP; P16649; 1ERJ.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 8.  
 DR PRINTS; PR00320; GPROTEINERPT.  
 DR PRODOM; PD000018; WD40; 4.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 2.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 674 AA; 76535 MW; E7521B469F8E80P7 CRC64;  
  
 Query Match 14.2%; Score 98.5; DB 2; Length 674;  
 Best Local Similarity 24.4%; Pred. No. 43;  
 Matches 30; Conservative 22; Mismatches 44; Indels 27; Gaps 5;  
  
 QY 15 LPVYKGELEKGYQFDGHEISGFEGKDGAGVNLNLSKDTPIK-----PVPKKEE--- 63  
 DB 524 LAVFRGHVGPAYKI-AWSI-----DNNYIISCSQDSTLKLWRINHLVPLKKCBENG 575  
 QY 64 -----KKKEENKPTPDV--SKKDNPOVNHSQLNESHRKEDLQREHSQKSDSTKVTV 115  
 DB 576 QPKSDPRNDQKNDQKDKDKDDQKNDKNDHANNQEDGEEKKKKEKXDKIKSKIK 635  
 QY 116 TVL 118  
 DB 636 TLL 638  
  
 RESULT 40  
 Q55G46 DICDI PRELIMINARY; PRT; 827 AA.  
 AC Q55G46;  
 DT 13-SEP-2005 (TReMBLrel. 31, Created)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=DOB0189573;  
 OS Dictyostelium discoideum (slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AX4;  
 RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,  
 RA Sugrang R., Berriman M., Song J., Olsen R., Szafranski K., Xu O.,  
 RA Tungal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,  
 RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
 RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,  
 RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
 RA Fabrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,  
 RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
 RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
 RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,  
 RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
 RA Loulseg H., Mungall K., Oliver K., Price C., Quail M.A.,  
 RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,  
 RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
 RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,  
 RA Shaubky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,  
 RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,  
 RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,  
 RT "The genome of the social amoeba Dictyostelium discoideum.";  
 RL Nature 0:0-0(2005).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.



DR EMBL; AAFI01000009; EAL73367.1; -; Genomic DNA.

**KW** Hypothetical protein; Nucleotide-binding. [EMBL:U00000.1](#); [GenBank:U00000.1](#); [UniProt:U00000.1](#)

SQ SEQUENCE 827 AA; 92700 MW; D576462666E4469 CRC64; hyponeuritic protein, nucleotide-binding.

Query Match 14.2%; Score 98.5; DB 2; Length 827;  
Best Local Similarity 26.9%; Pred. No. 53;  
Matches 29; Conservative 17; Mismatches 31; Indels 31; Gaps 4;

QY 57 VFKIEKK--EENKPTFDVSKKKNPQVNHSQLNE--SHRKEDLQREEH----- 103

[illegible]

Qy	104	-----SQKSDSTKDV-----ATVLDKNNISSKSTNNPN	133
----	-----	--	-----

[illegible]

Search completed: April 24, 2006, 14:59:34

Job time : 83.3893 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 14.0895 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSSSTVSEEDFILPVYKG.....ATVLDKNNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- PIR\_80:\*
- 1: PIR1:\*
- 2: PIR2:\*
- 3: PIR3:\*
- 4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	696	100.0	2140	2 P95074	serine proteinase,
2	693	99.6	2144	2 A97942	metalloproteinase,
3	110	15.8	558	2 T18467	hypothetical prote
4	108.5	15.6	1038	2 J05497	claustrin - chicke
5	99	14.2	211	2 T25911	hypothetical prote
6	98	14.1	665	2 B71609	hypothetical prote
7	97.5	14.0	1345	2 S46817	hypothetical prote
8	95	13.6	348	2 I37271	cylicin II - human
9	95	13.6	622	2 A90570	lipoprotein (impor
10	94.5	13.6	312	2 G81339	probable membrane
11	94	13.5	210	2 T28771	hypothetical prote
12	94	13.5	535	2 T37189	hypothetical prote
13	93	13.4	1397	2 T10466	DNA topoisomerase
14	92.5	13.3	219	2 B72291	hypothetical prote
15	92.5	13.3	325	2 T18283	hypothetical prote
16	91	13.1	253	2 T32879	hypothetical prote
17	89.5	12.9	508	2 B81594	hypothetical prote
18	89.5	12.9	508	2 B86549	hypothetical prote
19	89.5	12.9	508	2 C72074	hypothetical prote
20	89	12.8	528	2 B96795	unknown protein P2
21	89	12.8	1888	2 T39009	hypothetical prote
22	89	12.8	3724	2 T18427	hypothetical prote
23	88.5	12.7	301	2 T33068	hypothetical prote
24	88.5	12.7	371	2 A71683	hypothetical prote
25	88.5	12.7	385	2 T20410	hypothetical prote
26	88.5	12.7	540	2 D86432	hypothetical prote
27	88.5	12.7	644	2 T47835	hypothetical prote
28	88.5	12.7	762	2 G88436	protein 104A9.13 [
29	88.5	12.7	791	2 T24435	hypothetical prote

30	88.5	12.7	2401	2 T28676	rhoptyr protein -
31	88	12.6	276	2 T23451	hypothetical prote
32	88	12.6	700	2 S67610	probable membrane
33	87.5	12.6	1702	2 A41859	IGA-specific metal
34	87.5	12.6	2523	2 T18477	hypothetical prote
35	87	12.5	635	2 P71621	hypothetical prote
36	87	12.5	2464	1 Q8MSF1	microtubule-associ
37	86.5	12.4	670	2 T28391	ORF MSV230 hypothe
38	86.5	12.4	792	2 T49989	hypothetical prote
39	86.5	12.4	2485	1 H71621	serine/threonine-s
40	86	12.4	1016	2 T19006	ankyrin related pr
41	85	12.2	335	2 T33457	hypothetical prote
42	85	12.2	867	2 T27136	hypothetical prote
43	85	12.2	871	2 T27135	hypothetical prote
44	85	12.2	988	2 T14188	hypothetical prote
45	85	12.2	1166	2 H71609	hypothetical prote

ALIGNMENTS

RESULT 1

P95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074

R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hea  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
neon, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrisor

A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: F95074

A;Status: preliminary

A;Molecule type: DNA

A;Cross-references: 1-2140 <KUR>

A;Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP0641

Query Match 100.0%; Score 696; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4.3e-46;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSSSTVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGVYINLSKDTFKPVPKK 60

Db 1973 KEMSSSTVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGVYINLSKDTFKPVPKK 2032

Qy 61 IEKKKEENKPTPDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 120

Db 2033 IEKKKEENKPTPDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 2092

Qy 121 NNISSSKSTTNNPNK 134

Db 2093 NNISSSKSTTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, I.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234







Nature 399, 323-329, 1999  
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A/Reference number: A72200; MUID:99287316; PMID:10360571  
A/Accession: B72291  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-219 <ARN>  
A/Cross-references: UNIPROT:Q9XOM6; UNIPARC:UPI00000C12ED; GB:AE001771; GB:AE000512; NID  
A/Experimental source: strain MSB8  
C/Genetics:  
A/Gene: TM1142

Query Match 13.3%; Score 92.5; DB 2; Length 219;  
Best Local Similarity 28.1%; Pred. No. 2.8;  
Matches 36; Conservative 24; Mismatches 35; Indels 33; Gaps 8;  
QY 20 GLEKGYO--FGWISG-----PEGKKDAGYVIN-LSKOTFIPVFKIEKKENK 70  
DB 89 GSESVYRDLFAGFVGRTPTFPFPGKEGLGLPGYVDKNFIK-ILKYVAQELKED-- 145  
QY 71 PTFDVSKKKQNPQVNHSQLNESHKED---LQREHSQKSDS-----TKD 112  
DB 146 --FQYLYKKDDPFVGEPLIIEIF-KEDADFVLEKDENAVKDTPVNEVRDRIRYVTDSPD 202  
QY 113 VTATVLDK 120  
DB 203 VAKTLQEK 210

RESULT 15  
B18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C/Species: Dictyostelium discoideum  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
A/Accession: T18283  
R/Riechen Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hught  
Genetics 148, 1117-1125, 1998  
A>Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A/Reference number: Z14684; MUID:98198836; PMID:9539429  
A/Accession: T18283  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-325 <RIB>  
A/Cross-references: UNIPROT:Q44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:G2702254; PI  
C/Genetics:  
A/Introns: 85/1

Query Match 13.3%; Score 92.5; DB 2; Length 325;  
Best Local Similarity 23.2%; Pred. No. 4.3;  
Matches 38; Conservative 23; Mismatches 42; Indels 61; Gaps 8;  
QY 8 VSEEDFILPVVK-GEL--EKGYOPDGWISGPEGK-----KDAGYV-----I 46  
DB 72 IGDEKELFRISKNGELIVLNELEPNFIK--EGHLKRSKMFNHIKSGYATNNEETEI 129  
QY 47 NLSKDTFIKPV-----FKK-----IBKKKEENKPTFDV 75  
DB 130 FLESCTLCKBITAQTAKRNSYKRNINKLPBEEEEEEEEEEEEEEEEEEEEEEQEEVEKPTISE 189  
QY 76 SKKKNPQVNHSQLNESHKED-----DLQREHSQKSDTKD 112  
DB 190 EEEETPAVSEBEKEEEEEETPAVSEBEKEEEEEQEDKEKD 233

RESULT 16  
T32879  
hypothetical protein C17F3.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
A/Accession: T32879  
R/Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A/Description: The sequence of C. elegans cosmid C17F3.

A/Reference number: Z21240  
A/Accession: T32879  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-253 <GAT>  
A/Cross-references: UNIPROT:Q44948; UNIPARC:UPI0000074BB9; EMBL:AF043692; PIDN:AAB97531  
A/Experimental source: strain Bristol N2; clone C17F3  
C/Genetics:  
A/Gene: CESP:C17F3.3  
A/Map position: 1  
A/Introns: 41/1

Query Match 13.1%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 4.2;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;  
QY 56 PVFKKIEKKK--EENKPTFDVSKK-----KDNQVNHSQLNESHK----- 94  
DB 61 PVAPKVEKKKEKKKEKKADDEKKTEKKDKKSKKTEKKDKKISVKKTKTQTKSERKKK 120  
QY 95 ---REDLQREHSQKSDTKDVTATVLDKNNISSK 126  
DB 121 DEKKEDKKKSKSKSKDEKKKDEKDKKDEKK 155

RESULT 17  
B81594  
hypothetical protein CP0281 [imported] - Chlamydomophila pneumoniae (strain AR39)  
C/Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
A/Accession: B81594  
R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzber  
Nucleic Acids Res. 28, 1397-1406, 2000  
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A/Reference number: A81500; MUID:20150255; PMID:10684935  
A/Accession: B81594  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-508 <REA>  
A/Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:AE002189; GB:AE002161; N  
A/Experimental source: strain AR39, HL cells  
C/Genetics:  
A/Gene: CP0281

Query Match 12.9%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 12;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 54 IKPVFKKIEKKKEENKPTFD-----VSKKKDNQVNHSQLNESHK 95  
DB 92 VKGVFKKTPQARPEVSSPRLPSHVHGQRLPGLRGFRDRIOKRSENPEADLGKMKRSYSD 151  
QY 96 EDLQREHSQKSDTKDVTATVLDKNNISSKTT 129  
DB 152 GDLDRVGHDSNEDSTEDSRS---EGGEPSSKSS 182

RESULT 18  
B86549  
hypothetical protein CPJ0473 [imported] - Chlamydomophila pneumoniae (strain J138)  
C/Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
A/Accession: B86549  
R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; I  
Nucleic Acids Res. 28, 2311-2314, 2000  
A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A/Reference number: A86491; MUID:20330349; PMID:10871362  
A/Accession: B86549  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-508 <STO>  
A/Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:BA000008; NID:G8978843; I





Query Match 12.8%; Score 89; DB 2; Length 3724;  
Best Local Similarity 22.3%; Pred. No. 1.2e+02;  
Matches 29; Conservative 27; Mismatches 42; Indels 32; Gaps 5;  
  
QY 6 TIVSEEDPILPVYKGELEKGYQPDGWEISGFGKKDAGVYNLSKDTFIKPVFKIERKK 65  
DB 1060 TWICENNIEQNEENSKGVISGTDN---ENKND-----MENKN 1097  
QY 66 EENKPTFDVSKKKNPQVNH-SQLNESHKEDLQREHSQKSDTKQVATVLDKNNIS 124  
DB 1098 DMEKKN--DMEKKNDIEKKNDMEKKNDMEKKNDMEKKNDME-----NENNWE 1148  
QY 125 SEKSTNNPNK 134  
DB 1149 NKSDIENENK 1158

## RESULT 23

T33068  
Hypothetical protein C35E7.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T33068  
R:Graves, T.; McDonald, R.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid C35E7.  
A:Reference number: Z21278  
A:Accession: T33068  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-301 <GRA>  
A:Cross-references: UNIPROT:O61765; UNIPARC:UPI0000074979; EMBL:AF067216; PIDN:AAC17524.  
A:Experimental source: strain Bristol N2; clone C35E7  
C:Genetics:  
A:Gene: CESP:C35E7.9  
A:Map position: 1  
A:Introns: 30/3; 193/1; 236/2

Query Match 12.7%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 8.1;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;  
  
QY 25 GYQPDGWEISGEGKK-----DAGVYNLSKDTFIKPVFKI-----EB 63  
DB 10 GITIAGWILAGCGKXKXGKSTASAAAPKADSKKPPVENVSKSKSEKKEPKKEBP 69  
QY 64 KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKD 112  
DB 70 KKEBKKEKSKSEKKDKK-----EEAKGDDKDKKEDKEDKDKD 112

## RESULT 24

A71683  
Hypothetical protein RP278 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: A71683  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71683  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-371 <AND>  
A:Cross-references: UNIPROT:Q9ZDP9; UNIPARC:UPI00000C1161; GB:AJ235269; NID  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: RP278

Query Match 12.7%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 10;

Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;  
QY 1 KEMSTIVSEBDPILPVYKGELEKGYQPDGWEISGEGKK-DAGVYNLSKDTFIKPVFK 59  
DB 192 EQLNKTILNEB-----FRKNLEQ-----EKKKIEGKKTQASEALNKK----LKPIYK 234  
QY 60 KIEBKKEB-----ENKPTFDVSKKKNPQVNH---HSQLNESHKEDLQREHSQKSDTK 111  
DB 235 QMDEREELFKLAELTPOY-----AQAINDHAKLYAKQYQTKIENDPNYKLEKLG 286  
QY 112 DVTATVLDKNNISSKSTT-----NNPN 133  
DB 287 E-----IVSKIEYKSKSTNDIINNPN 309

## RESULT 25

T20410  
Hypothetical protein E02A10.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20410  
R:Thomas, K.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19271  
A:Accession: T20410  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-385 <WIL>  
A:Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z81053; PIDN:CAB02877.1  
A:Experimental source: clone E02A10  
C:Genetics:  
A:Gene: CESP:E02A10.2  
A:Map position: 5  
A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.7%; Score 88.5; DB 2; Length 385;  
Best Local Similarity 35.5%; Pred. No. 11;  
Matches 22; Conservative 14; Mismatches 19; Indels 7; Gaps 2;

QY 59 KKEBKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKQVATV 118  
DB 317 KKEBKKEEKEQKE--EVEKKEE-----EKKDKEPKKEEKEEKEEKEEKESEK 369  
QY 119 DK 120  
DB 370 EK 371

## RESULT 26

DB6432  
Hypothetical protein T518.14 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: DB6432  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziah,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: DB6432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>  
A:Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI00000ABF35; GB:AE005172; NID:G4587525; I

```

Query Match          12.7%; Score 88.5; DB 2; Length 540;
Best Local Similarity 24.4%; Pred. No. 15;
Matches 31; Conservative 25; Mismatches 54; Indels 17; Gaps 4;

QY  2 EMSTVISEEDFILPVYKGLKGYQDFGWEISGFEGKQAGYVNLSTKTFIKPVFKK- 60
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  49 EKASPFKEESDFFADLKESEKK-----ALSLKSLKEAIVDN---TLTKKKES 96
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  61 --IEEKKEENKPTFDVSKKONQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVL 118
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  97 SPNKEKKEEVVYKPAEVEKKKE--EAAAEKVEEKKSEAVVTEEPKAAETVEAVVTEII 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  119 DKNNISS 125
    | : : :
Db  155 PKEEVT 161

RESULT 27
T47835
hypothetical protein T209.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
R;Accession: T47835
R;Nykatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24475
A;Accession: T47835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-644 <NVA>
A;Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658
A;Experimental source: cultivar Columbia; BAC clone T209
C;Genetics:
A;Map position: 3
A;Introns: 158/2; 329/3
A;Note: T209.90

```

Query Match	12.7%; Score 88.5; DB 2; Length 644;
Best Local Similarity	24.6%; Pred. No. 19;
Matches 32; Conservative 18; Mismatches 53; Indels 27; Gaps 3;	
Qy	9 SEEDFLLPVVKGELEKGVQPDGWEISFGEGKKDAGVYINLSKDTPIKVPVKIEEKEEE 68     : : :     : :     : :     : :     : :
Db	534 SVADFLLRIKKNSPQKG-----ETTSKNQKNDGNV-----KKEND 570     : : :     : :     : :     : :     : :
Qy	69 NKPTFFOVSKKDNQVNHSQLNESHRKEDLQREHSSOKSDS----TKDVTATVLDDKNNIS 124 : :     : :     : :     : :     : :     : :
Db	571 HQKSGDNVKKENSKVAPRELRLSRSTGGKKVEVENNSSKSKRKQTETAETAVATGCKGRE 630 : :     : :     : :     : :     : :     : :
Qy	125 SKSTTNPNK 134 
Db	631 SGKDKQP RK 640 

RESULT 28  
G88436  
protein T04A8.13 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: G88436  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes  
A;Reference number: A75000; PMID:98059613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999.  
A;Accession: G88436  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-762 <STO>  
A;Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000017A5C4; GB:chr\_III; PIDN:CAA84732.1;  
C;Genetics:  
A;Gene: T04A8.13

[illegible]

RESULT 29  
T24435  
hypothetical protein T04A8.13 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PDB:1C8A  
A:Experimental source: clone T04A8

A;map position: 2									
A;introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2									
Query Match 12.7%; Score 88.5; DB 2; Length 791;									
Best Local Similarity 25.2%; Pred. No. 24;									
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3									
Qy	9	SEEDPILPVYKGELEKCYQFDGWEISGFEGK-XDAGVYINLSKDTFFKPVFKKIEEKEE	67						
	:	:	:	:	:	:	:	:	:
Db	14	SGQIMKPGY--DKKEGLGMDQKEIVGDDKKDKARKRERKLQDEFAB--LKDEEEKDKE	69						
	:	:	:	:	:	:	:	:	:
Qy	68	ENKPTPDVSKCONPVNHSQLNESHREKDLOREHSQKSDSTKDVATVLDKNNISSKS	127						
	:	:	:	:	:	:	:	:	:
Db	70	EAEKKEKKEKGEKEEKEDGHEKKEDKEDKENENDEKKEKSDDKKESKEDKKEKTK	129						
	:	:	:	:	:	:	:	:	:
Qy	128	TTNNPNK	134						
	:	:	:	:	:	:	:	:	:
Db	130	TEDNEGK	136						
	:	:	:	:	:	:	:	:	:

```

RESULT 30
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A4521
C:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q6216; UNIPARC:UPT000017B647; EMBL:U36927; NID:g1041784; P:
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990

```



A;Map position: 3  
A;Introns: 148/3  
A;Note: C0485w

Query Match	12.6%	Score 87.5;	DB 2;	Length 2523;
Best Local Similarity	32.5%;	Pred. No. 1e+02;		
Matches 25;	Conservative 22;	Mismatches 21;	Indels 9;	Gaps 5;
Qy	59	KKIEEKEEENKPTFDVSKKKD-NPQVNHSLQNSHREKL-QREHSQKSDSTK--DVT	114	
Db	1795	KKKKIKKKDDINKKE-DINKKODINKKONINKKODINKKODINKKODINKKODINKKODINK	1853	
Qy	115	ATVLDKNNISKSKSTTN	131	
Db	1854	-----KKNNYNNNNNNNN	1866	

RESULT 35  
F71621  
hypothetical protein PFB0170w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: F71621  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: F71621  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-635 <GAR>  
A:Cross-references: UNIPROT:O96137; UNIPARC:UPI000007E196; GB:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0170w

	Query Match	12.5%	Score 87;	DB 2;	Length 635;
	Best Local Similarity	23.8%	Pred. NO. 24;		
	Matches 30;	Conservative 20;	Mismatches 52;	Indels 24;	Gaps 4;
Qy	22	LEKGYQDGEISGFEKGDAGYINLSKOTPIKPVFKTIEEKEENKTFDVSKKDN	81		
Db	202	LQKYNITQDDEEDNETIRSDKLIDYSQSKDIMSSSPNKEEES-----MSSDNHN	256		
Qy	82	PQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLD-----KNNISKS	127		
Db	257	KDINSS-----DNQNKDINSSDHNW-NDSNTESTTSLSTSIINTNRNKKKKNININN	311		
Qy	128	TTNNPN	133		
Db	312	NNNNNN	317		

RESULT 36  
QRMSP1  
microtubule-associated protein MAP1B - mouse  
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 09-Jul-2004  
C:Accession: S07549; S44387; A33645  
R:Noble, M.; Lewis, S.A.; Cowan, N.J.  
J. Cell Biol. 109, 3367-3376, 1989  
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains  
A:Reference number: A33645; MUID:90094539; PMID:2480963  
A:Accession: S07549  
A:Molecule type: mRNA  
A:Residues: 1-2464 <NOB>  
A:Cross-references: UNIPROT:P14873; UNIPARC:UPI00000297D7; EMBL:X51396; NID:g52999; PIDN  
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.  
Arch. Biochem. Biophys. 310, 428-432, 1994  
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.  
A:Reference number: S44387; MUID:94234720; PMID:8179328

A;Accession: S44387  
A;Status: preliminary  
A:Molecule type: Protein  
A:Residues: 653-663,'IC' <SAN>  
A:Cross-references: UNIPARC:UIP0000173D97  
C;Superfamily: microtubule-associated p:protein MAP1B  
F;Keywords: Microtubule binding; phosphoprotein; tandem repeat  
F;S59-786/Domain: microtubule binding #status experimental <MTB>  
R-K-E/D-X)  
F;589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-699

F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

P;1961-2064/Region: 17-residue repeats  
F;94,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: phos  
F;147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co  
F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match            12.5%; Score 87; DB 1; Length 2464;  
Best Local Similarity     28.4%; Pred. No. 1.le+02;  
Matches      29; Conservative       16; Mismatches    37; Indels     20; Gaps     4;

Qy          48 LSKDTFIKVPFK-IIEKKKEEENKPTFDVSKKKDPQNVSQSLSHRKEDLQRE----- 101  
             :|||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
Db          632 VTOKVVVKKEIKTKLEEKKEG--KPKEKVWKCKXTPL---KKDEKPRKEEVKEIKEI 686  
  
Qy          102 -----EHSQSDSTKDVTATVLDKNNISKSSTTNPNK 134  
             :|||:|||  
Db          687 KKERKELKGVKKETPLDKAKVEVKGEKVEVKKEEPKK 728

RESULT 37  
T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C;Species: Melanoplus sanguinipes entomopoxvirus  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T28391  
F;Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A;Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A;Reference number: Z20484; PMID:99102612; PMID:9847359  
A;Accession: T28391  
A;Status: preliminary; translated from GB/EMBL/DDBB  
A:Molecule type: DNA  
A;Residues: 1-670 <AF0>  
A:Cross-references: UNIPROT:Q9YVL2; UNIFARC:UIP00000F2D3E; EMBL:AF063866; NID:g4049647;  
C;Genetics:  
A>Note: MSV230

Query Match            12.4%; Score 86.5; DB 2; Length 670;  
Best Local Similarity     22.7%; Pred. No. 28;  
Matches      35; Conservative       33; Mismatches    65; Indels     21; Gaps     5;

Qy          1 KEMSSTI-----VSEDFILPVYKGELEGY----QFDGWGISGFEGKKDA 42  
             |||::|||::|||::|||::|||::|||::|||::|||::|||:  
Db          39 KOILTLLKFNFDMCGSVSEVFQLINKNSASEKYSDVDSSIDESQNSDSDDS 98  
  
Qy          43 GVINLSKDTPIKVPFKIB-EKKEENKPTPDVSKKKDNQPVNHSQL-NESHREDLQR 100  
             |||::|||::|||::|||::|||::|||::|||::|||::|||:  
Db          99 GWNIDESQNSDKVNINKLENESQNSDSKNVIDEQNSDS-KVNINKLENESQNSDSKV 157

Qy          101 EEHQSDSTKDVTATVLDKNNISKSSTTNPNK 134  
             :|||||::|||::|||::|||::|||::|||::|||::|||:  
Db          158 IDESQNSDKNVINDESQNSDSKNVIDEQNSDSK 191

RESULT 38  
T49989  
hypothetical protein F12b17.150 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
C;Accession: T49989  
F;Bevan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence database, April 2000  
A;Reference number: Z25026  
A;Accession: T49989



**THIS PAGE BLANK (w/09)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 10.4922 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEEDFILPVYKG.....ATVLDKNKISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA\_New:\*

- 1: /SID95/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /SID95/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /SID95/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /SID95/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /SID95/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /SID95/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /SID95/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /SID95/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	15.4	354	7	US-11-189-817-2
2	103.5	14.9	651	7	US-11-128-660-1
3	88	12.6	746	6	US-10-793-626-652
4	86	12.4	443	7	US-11-188-298-1015
5	85	12.2	472	6	US-10-793-626-658
6	85	12.2	720	6	US-10-793-626-2058
7	84.5	12.1	140	7	US-11-096-568A-4771
8	84.5	12.1	501	6	US-10-485-517-381
9	84.5	12.1	645	6	US-10-485-517-244
10	84.5	12.1	1694	7	US-11-052-554A-83
11	82	11.8	439	7	US-11-188-298-15964
12	82	11.8	439	7	US-11-188-298-16606
13	82	11.8	886	7	US-11-087-099-11456
14	81.5	11.7	8746	7	US-11-096-686-10232
15	79.5	11.4	493	7	US-11-096-568A-3070
16	79.5	11.4	493	7	US-11-096-568A-3071
17	79.5	11.4	510	7	US-11-096-568A-3069
18	79.5	11.4	943	6	US-10-475-204-34
19	79.5	11.4	1036	7	US-11-096-568A-28315
20	79.5	11.4	1070	7	US-11-096-568A-28314
21	79.5	11.4	1276	7	US-11-096-568A-28313
22	79	11.4	732	7	US-11-124-368A-248
23	79	11.4	803	7	US-11-124-368A-241
24	79	11.4	803	7	US-11-124-368A-242
25	78.5	11.3	258	7	US-11-098-686-10475

26	78.5	11.3	700	7	US-11-196-475-74	Sequence 74, Appl
27	78.5	11.3	1155	6	US-10-793-626-1780	Sequence 1780, Ap
28	78	11.2	191	6	US-10-860-601-5	Sequence 5, Appl
29	77.5	11.1	405	6	US-10-485-517-239	Sequence 239, App
30	77.5	11.1	627	6	US-10-873-528-191	Sequence 191, App
31	77	11.1	568	6	US-10-793-626-2482	Sequence 2482, Ap
32	77	11.1	663	7	US-11-196-475-70	Sequence 70, Appl
33	77	11.1	663	7	US-11-196-475-78	Sequence 78, Appl
34	77	11.1	1145	6	US-10-793-626-1432	Sequence 1432, Ap
35	76.5	11.0	439	7	US-11-188-298-17915	Sequence 17915, A
36	76	10.9	244	7	US-11-096-568A-11931	Sequence 11931, A
37	76	10.9	482	7	US-11-232-440-33	Sequence 33, Appl
38	76	10.9	502	7	US-11-045-004-1651	Sequence 1651, Ap
39	75.5	10.8	278	6	US-10-793-626-2700	Sequence 2700, Ap
40	75.5	10.8	404	6	US-10-793-626-398	Sequence 398, App
41	75.5	10.8	558	6	US-10-510-903-22	Sequence 22, Appl
42	75	10.8	513	7	US-11-185-924-16	Sequence 16, Appl
43	74.5	10.7	313	7	US-11-096-568A-32043	Sequence 32043, A
44	74.5	10.7	425	7	US-11-096-568A-32042	Sequence 32042, A
45	74.5	10.7	434	7	US-11-096-568A-32041	Sequence 32041, A

#### ALIGNMENTS

RESULT 1  
US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US2006030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM  
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES  
; FILE REFERENCE: 275601US0  
; CURRENT APPLICATION NUMBER: US/11/189,817  
; PRIOR FILING DATE: 2005-07-27  
; PRIOR APPLICATION NUMBER: 60/598,062  
; PRIOR FILING DATE: 2004-08-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-11-189-817-2

Query Match 15.4%; Score 107; DB 7; Length 354;  
Best Local Similarity 23.8%; Pred. No. 0.047;  
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
QY 1 KEMSTIVSEEDFILPVYKGLEKGYQPDGWEISGF--EGKODAG-----TVINLSKDTF 53  
Db 185 KEASS-----YDYIL-----GWFQGGVPEHKKEKRNMLSHLYVSSKDKENI 225  
QY 54 IKPVFKIEEKKEE-----ENKPTFDVSKKKONQVNHSHOLNESHKEDLQR 100  
Db 226 SKENDVDLDEKEAEETEEELKEKKEETESISEDEEEEEEKEEENDKKQEK 285  
QY 101 EHSQKSDSTKDVATVLDKNKISSKSTNN 131  
Db 286 EQSNENNDQKKDMEA-----QNLISKQNNN 311

#### RESULT 2

US-11-128-660-1  
; Sequence 1, Application US/11128660  
; Publication No. US2006002432A1  
; GENERAL INFORMATION:  
; APPLICANT: Statens Serum Institut  
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plas  
; TITLE OF INVENTION: falciparum  
; FILE REFERENCE: 15007dk



```
/ CURRENT APPLICATION NUMBER: US/11/128,660
/ CURRENT FILING DATE: 2005-05-12
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 651
/ TYPE: PRT
/ ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      14.9%; Score 103.5; DB 7; Length 651;
Best Local Similarity 25.2%; Pred. No. 0.19;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY      1 KEMSTIVSEEDPILPVVYKGELEKGYQDGEISGF--EGKKDAG-----YVINLSKDTF 53
DB      483 KEASS-----YDIIL-----GWFFGGVPEHKKEENMLSHLYVSSKKKENI 523

QY      54 IKPVFPKIEEKEEENKPTFDVSKKQNPQVNHSQLN-----BSHRKEDLQR 100
DB      524 SKENDVDLDE--KEEAEATEELEEKEEETESISEDEEEEEEEKEEENEKKEQEK 582

QY      101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB      583 EQSNENNQKKDMEA-----QNLISKQNNN 608

RESULT 3
US-10-793-626-652
/ Sequence 652, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 652
/ LENGTH: 746
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-652

Query Match      12.6%; Score 88; DB 6; Length 746;
Best Local Similarity 25.7%; Pred. No. 5;
Matches 37; Conservative 21; Mismatches 50; Indels 36; Gaps 7;

QY      4 SSTIVSEEDPILPVVYKGELEKGYQDGW---EISGF-----CKDAGYVIN--LSKD 51
DB      611 SGTKVLPHSKVMLMTDGLTNP-DMTGWTKEVDLAFELDTKLKVTGKNGFTWQSISKG 669

QY      52 TPIKVPFKIEEKEEENKPTDVS-----KKQNPQVNHSQLNESHRKEDLQREHSQKS 107
DB      670 QIIK-----NKKIEVLSAEDTDDQKTDSDSDNKKDKADEDSNNTS 716

QY      108 DSTKDVATVLDKNNISSKSTNN 131
DB      717 SSTKN-----DKSNADSKNDSDD 734

RESULT 4
US-11-188-298-1015
/ Sequence 1015, Application US/11188298
/ Publication No. US20060075522A1
/ GENERAL INFORMATION:
/ APPLICANT: Abad, Mark S. et al.

/ TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
/ FILE REFERENCE: 38-21(53452)B
/ CURRENT APPLICATION NUMBER: US/11/188,298
/ CURRENT FILING DATE: 2005-07-22
/ PRIOR APPLICATION NUMBER: 60/592,978
/ PRIOR FILING DATE: 2004-07-31
/ NUMBER OF SEQ ID NOS: 22569
/ SEQ ID NO 1015
/ LENGTH: 443
/ TYPE: PRT
/ ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match      12.4%; Score 86; DB 7; Length 443;
Best Local Similarity 17.8%; Pred. No. 4.1;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY      18 YKGELEKGYQDGEISGFEGKKDAGYVINLSKDTFI----- 54
DB      42 YEEAVEDGVSGFDGSSIPGFGIEDSLIPKADPSTYAEIPWEGIGRVGVYIKGDEPYQA 101

QY      55 --KVPFKIEEKEEEN-----KPTFDVSKKKD----- 80
DB      102 DPRGILKRVLERLEKEGLKAHIGPEFYIFKQGTWELHIPDSGGYFDLVGLDKAREIR 161

QY      81 -----NPOVNHSQLNESHRKEDLQREHSQKSD---STKDVATVLD 119
DB      162 REIALYMPYLGLKPEVLHHEVGKAQHEIDFRYDEALTAADNIVSFKHVVKAAVE 215

RESULT 5
US-10-793-626-658
/ Sequence 658, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 658
/ LENGTH: 472
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-658

Query Match      12.2%; Score 85; DB 6; Length 472;
Best Local Similarity 32.6%; Pred. No. 5.4;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY      61 IEKKEEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQ-----REEHSQKSDS--- 109
DB      3 MEENKQNPKE--NMSNKDDNA-----TLNDSHREDELELFRNNKNARQRRRRIDNQSK 56

QY      110 TKDVAT-----VLDKNNISSKSTNNPK 134
DB      57 EKDATSTQSQLETKPMDFLDNHNKS--HNQNK 86

RESULT 6
US-10-793-626-2058
/ Sequence 2058, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
```

```
; FILE REFERENCE: PUB480US
; CURRENT APPLICATION NUMBER: US/10/793.626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match          12.1%; Score 85; DB 6; Length 720;
Best Local Similarity 28.0%; Pred. No. 8.8;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 61 IEEKKEENKPTPDVSKKD----NPQVNHSQLNESHKEDLQREHHSQKSDSTK----D 112
      :|||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 38 LEEQIKALDKFKASQAKDTNKQNTQNNHQSKNNKQNSNDKEKQSQKNNKPTKKKEQN 97

Qy 113 VTATVLDKNNISSKSTTNNPNK 134
      :|||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 NKGQKQNNKNTNKQNNKNNK 119

RESULT 7
US-11-096-568A-4771
; Sequence 4771, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096.568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4771
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(140)
; OTHER INFORMATION: Ceres Seq. ID no. 14304111
US-11-096-568A-4771

Query Match          12.1%; Score 84.5; DB 7; Length 140;
Best Local Similarity 26.6%; Pred. No. 1.4;
Matches 33; Conservative 25; Mismatches 39; Indels 27; Gaps 6;

Qy 10 BEDFILPVY--KGELEKGYQDQWEGISGFEGKQDAGYVINLSKDTFKPVFKIEKKKE 67
      |||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 EDEKIPAYRRGRGPQPMKD-----FEEDEE-----EELVEKKEEEEE 65

Qy 68 ENKPTFDVSKKQNPQ-----VNHSQLNESHKEDLQREHHSQKSDSTKQVATVLDKNNI 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 DDSVT---SKKEBERKFKMINGSNTDANEKENGIGSKSRDGSSTKS-TSTGFRQNGS 121

Qy 124 SSKS 127
      ||
Db 122 RRKS 125

RESULT 8
US-10-485-517-381
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
```

```
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485.517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match          12.1%; Score 84.5; DB 6; Length 501;
Best Local Similarity 28.6%; Pred. No. 6.4;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 41 DAGYVINL-SKDTFIKPVFKIEKKEENKPTFDV-----SKKQNPQVNHSQLNESH 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 DGQYHVRIVDKAEFTKANTDKSNKKEQODNSAKKEATPATPSKPTSPVKEKESQKQDSQK 356

Qy 95 KEDLQ----REHHSQKSDSTKQV--ATVLDKNNISSKSTTNNPNK 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 DDNKQLPSVEKENDASSESGDKTTPATKPTKGEVSSSTT--PTK 399

RESULT 9
US-10-485-517-244
; Sequence 244, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485.517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 244
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-244

Query Match          12.1%; Score 84.5; DB 6; Length 645;
Best Local Similarity 28.6%; Pred. No. 8.6;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 41 DAGYVINL-SKDTFIKPVFKIEKKEENKPTFDV-----SKKQNPQVNHSQLNESH 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 DGQYHVRIVDKAEFTKANTDKSNKKEQODNSAKKEATPATPSKPTSPVKEKESQKQDSQK 500

Qy 95 KEDLQ----REHHSQKSDSTKQV--ATVLDKNNISSKSTTNNPNK 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 DDNKQLPSVEKENDASSESGDKTTPATKPTKGEVSSSTT--PTK 543

RESULT 10
US-11-052-554A-83
```

```
; Sequence 83, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Haemophilus influenzae Rd
US-11-052-554A-83

Query Match      12.1%; Score 84.5; DB 7; Length 1694;
Best Local Similarity 26.1%; Pred. No. 27;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy      46  INLSKDTFKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS- 104
Db      1288  INTGATATETAEKSDKQFETASTEDASQHKANTVADNSVANNSESSDPKRRRSI 1347

Qy      105  --QKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db      1348  SQPQTSABETTAASTDETTIADNKRKSPNR 1379
```

```
RESULT 11
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; PRIOR FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964

Query Match      11.8%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 9;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy      25  GYQFGWEISGPEGKKGAGYVNLKDTFI-----KP-----VFK 59
Db      45  GISFDGSSVPGFGIEDSLVFKADEFTYVEVPWDMNVARVGYFYKONKPYGADPRGILK 104

Qy      60  KIEEKEEN-----KPTFDVSKKKN-----81
Db      105  RALELEKEGYKAYIGPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164

Qy      82  -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 123
Db      165  PSFGLIPEVLHVEVCKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215
```

```
RESULT 12
US-11-188-298-16606
; Sequence 16606, Application US/11188298
```

```
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; PRIOR FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 16606
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus wossei
US-11-188-298-16606

Query Match      11.8%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 9;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy      25  GYQFGWEISGPEGKKGAGYVNLKDTFI-----KP-----VFK 59
Db      45  GISFDGSSVPGFGIEDSLVFKADEFTYVEVPWDMNVARVGYFYKONKPYGADPRGILK 104

Qy      60  KIEEKEEN-----KPTFDVSKKKN-----81
Db      105  RALELEKEGYKAYIGPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164

Qy      82  -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 123
Db      165  PSFGLIPEVLHVEVCKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215
```

```
RESULT 13
US-11-087-099-11456
; Sequence 11456, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11456
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Bacillus cereus ATCC 14579
US-11-087-099-11456

Query Match      11.8%; Score 82; DB 7; Length 886;
Best Local Similarity 24.4%; Pred. No. 20;
Matches 40; Conservative 25; Mismatches 53; Indels 46; Gaps 8;

Qy      11  EDTFLPVY--KGEL-----EKGYQF--DGWEIS-GFEGKGD 41
Db      557  EEIQIPVYDLGESEIENQLVSEGGTFNNGVTKWSTFGEKYKFDLSDSEISIRNGI-- 614

Qy      42  AGYVINLAKDTFFIKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQ-- 99
Db      615  --VIQNIWEKEEKEPEVKIEETKEEKEPEVKIEETKEEVPKVEET--KEEVKEP 670

Qy      100  ---REEHSQKSDTKDVTATVLD-----KNNISSKSTNNPNK 134
Db      671  TKVEEAKKEVKPTKEEVEETKEEVPKVEEAKKEVEKEPTK 714
```

```
RESULT 14
US-11-098-686-10232
; Sequence 10232, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10232
; LENGTH: 8746
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10232

Query Match      11.7%; Score 81.5; DB 7; Length 8746;
Best Local Similarity 25.9%; Pred. No. 3.3e+02;
Matches 38; Conservative 24; Mismatches 60; Indels 25; Gaps 7;

Qy  2 EMSSTIVSEEDFIL--PVYKGELEKGYQFDGWE--ISGFEGKKGADGYYVNLKSDTFIKPV 57
Db  3358 EGEDVAEEEDLLWLDLNEAEGTKEEHAEVKVGEVGEVFDG-----ISEED--KP- 3410
Qy  58 FKKEEKEEENKPTFDVSKKKNPQVNHSQLNESHKEDL-----QREHESQKS 107
Db  3411 --KKDDKEQKATGLDGSGETIEESQPOQEKEEKENSFGSNESPSPOEESYDE 3468
Qy  108 DSTKDVATVLDKNNISSKSTTNNPNK 134
Db  3469 TSSVVTSPLLSINEV--KQTEDESAK 3493

RESULT 15
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

Query Match      11.4%; Score 79.5; DB 7; Length 493;
Best Local Similarity 25.5%; Pred. No. 17;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGYYI----- 46
Db  90 ANTIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSDDKDEGNVAVQAQDD 146
Qy  47 -----NLSKDTPIKVPFKIEB--KKEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  147 TQSTGKQTANADTTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 201
Qy  98 LQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db  202 DKKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 16
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3071
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match      11.4%; Score 79.5; DB 7; Length 493;
Best Local Similarity 25.5%; Pred. No. 17;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGYYI----- 46
Db  90 ANTIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSDDKDEGNVAVQAQDD 146
Qy  47 -----NLSKDTPIKVPFKIEB--KKEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  147 TQSTGKQTANADTTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 201
Qy  98 LQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db  202 DKKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 17
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match      11.4%; Score 79.5; DB 7; Length 510;
Best Local Similarity 25.5%; Pred. No. 18;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGYYI----- 46
Db  107 ANTIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSDDKDEGNVAVQAQDD 163
Qy  47 -----NLSKDTPIKVPFKIEB--KKEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  164 TQSTGKQTANADTTVTPEVKTKGVVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 218
```











```
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-78

Query Match          11.1%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 40;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

QY 55 KPVFKKIEEKKEENKPTFVSKKDNQVNHSHQ-LN-ESHKEDLQREHSQKSDSTKD 112
DB 378 REIEKQIEIKKND-----ELFKDKHAKLDLKQLNSKASSKEKIEGEERDKELDSKN 432

QY 113 V-TATVLDKNNISSKSTNNPNK 134
DB 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 34
US-10-793-626-1432
; Sequence 1432, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1432
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1432

Query Match          11.1%; Score 77; DB 6; Length 1145;
Best Local Similarity 24.8%; Pred. No. 75;
Matches 32; Conservative 22; Mismatches 41; Indels 34; Gaps 6;

QY 3 MSSTTVSEDFLTPVYKGELEKGYQFDGWISGFEKGKDGAGVYVNLSDTKTFIKPVFKKIE 62
DB 412 VNEKILNQEDYVPSLISEL---YQTQ-TSISN-----TYVL-----TFSDEVIKALN 455
```

```
QY 63 EKKEENKPTFVSKKDNQVNHSHQKEDLQREHSQKSDSTKDTATVLDKNN 122
DB 456 KCIENESTPLPEEA-----VNHVQVNELSSDENEDRYEYDR-----YIELNT 497

QY 123 ISSKSTNN 131
DB 498 LKDSILTSHN 506

RESULT 35
US-11-188-298-17915
; Sequence 17915, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/183,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 17915
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-11-188-298-17915

Query Match          11.0%; Score 76.5; DB 7; Length 439;
Best Local Similarity 17.5%; Pred. No. 27;
Matches 27; Conservative 23; Mismatches 35; Indels 69; Gaps 4;

QY 25 GYQPDGWEISGFEKGKDGAGVYVNLSDTKTFI-----KP-----VFK 59
DB 45 GISFDGSSVPGFGIEDSLVFKADPTTYVEVPMDNVARVYGVYKONKPYGADPRGILK 104

QY 60 KIEEKKEEN-----KPTFVSKKDN----- 81
DB 105 RALBELEKGEKGYKAYIGPEPEFYLFKNGTWELEIPDVGVPDILTLDKARDIRREIAEYM 164

QY 82 -----POVNHSHQKEDLQREHSQKSDS 109
DB 165 PSFGLIPEVLHVEVKAQHEIDFIDEALKTADN 198

RESULT 36
US-11-096-568A-11931
; Sequence 11931, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 11931
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(244)
; OTHER INFORMATION: Ceres Seq. ID no. 13659132
US-11-096-568A-11931

Query Match          10.9%; Score 76; DB 7; Length 244;
Best Local Similarity 26.4%; Pred. No. 15;
Matches 33; Conservative 21; Mismatches 41; Indels 30; Gaps 6;

QY 21 ELEKGYQPDGW---EISGFEKGKDGAGVYVNLSDTKTFIKPVFKKIEEKKEENKPTFVSK 77
```

```
Db 139 EKRRKIDGWRKQIASEAKENANFV-PLGGDWRDRVKKRABAKKEAKTEPIRAAA- 196
Qy 78 KKNDFQVNHSQLNESHKE-DLQREHSQKS-----DSTKDVATATVLDKNNISSKSTT 129
Db 197 -----EQHKGPDLSLSKGPFGSCQQAVIDESTKQ-----VYGGNNLTSETTW 239
Qy 130 NNPKN 134
Db 240 DRPSK 244

RESULT 37
US-11-232-440-33
; Sequence 33, Application US/11232440
; Publication No. US20060068434A1
; GENERAL INFORMATION:
; APPLICANT: STOECKER, JAY
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING CANCER USING
; TITLE OF INVENTION: COMPONENTS OF THE U2 SPLICEOSOMAL PARTICLE
; FILE REFERENCE: MTP-031
; CURRENT APPLICATION NUMBER: US/11/232,440
; PRIOR FILING DATE: 2005-09-21
; PRIOR FILING DATE: 60/612,310
; PRIOR FILING DATE: 2004-09-21
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: Patent In Ver. 3.3
; SEQ ID NO 33
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-232-440-33

Query Match 10.9%; Score 76; DB 7; Length 482;
Best Local Similarity 21.4%; Pred. No. 33;
Matches 27; Conservative 24; Mismatches 39; Indels 36; Gaps 5;

Qy 34 SGFEGKQAGYVINI-SKDTFKVPFKIIEKK-----EENKPTFDVSK 77
Db 40 SGLSQKEE-----EEDTFIEE--QLSEKLLERQRHLHEWLLREKQAQEFRIKK 90
Qy 78 KKNDFQVNH-----SQLNESHKEDLQREHSQKSDTKQVATVLD--KNNISSK 126
Db 91 EKBEAAKTRQBEQRKLKEQWEEQQRKEBEERQKQKKEBEALQMLDQAELENG 150
Qy 127 STNNP 132
Db 151 TTQNP 156

RESULT 38
US-11-045-004-1651
; Sequence 1651, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
; APPLICANT: BUCHRIEGER, CARMEN
; APPLICANT: FRANGEUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNIOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSSEURGET, OLIVIER
; APPLICANT: CHETOUANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GOEBEL, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
```

```
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOMANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARNIA
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUF, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01118
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: FR 00/04,629
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 2854
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 1651
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-11-045-004-1651

Query Match 10.9%; Score 76; DB 7; Length 502;
Best Local Similarity 25.5%; Pred. No. 35;
Matches 24; Conservative 17; Mismatches 25; Indels 28; Gaps 4;

Qy 10 BEDFILPVYKGE-LEKGYQFDGWEI-----SGFEGKKD-----AGY 44
Db 6 EEEFLAYIKKQEALEALALVYWDLTGAPAKGMEGRSDVIGLSEEIFNMQTSEMAAF 65
Qy 45 VINLSKDTFKVPFKIIEKKKEENKPTFDVSKK 78
Db 66 IAGLNQD----KENLSEITRKLTEESQKTYDLNKK 96

RESULT 39
US-10-793-626-2700
; Sequence 2700, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2700
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Artificial Sequence
```



GenCore version 5.1.7  
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 67.1499 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773  
Perfect score: 696  
Sequence: 1 KEMSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKQDAGYVNLKDTFIKPVFKK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	696	100.0	773	4	US-10-067-385-8
2	696	100.0	2119	3	US-09-769-744A-28
3	696	100.0	2140	4	US-10-282-122A-73670
4	696	100.0	2140	5	US-10-472-928-1180
5	693	99.6	637	5	US-10-617-320-3169
6	615	88.4	117	3	US-09-765-272-68
7	615	88.4	117	6	US-11-106-649-68
8	107	15.4	188	5	US-10-691-672A-7
9	103.5	14.9	169	5	US-10-691-672A-2
10	103.5	14.9	647	5	US-10-691-672A-3
11	101.5	14.6	564	6	US-11-097-143-12723
12	99.5	14.3	707	4	US-10-282-122A-52942
13	98	14.1	665	3	US-09-820-843A-107
14	93	13.4	869	4	US-10-437-963-12282
15	93	13.4	1529	5	US-10-732-923-22588
16	93	13.4	3127	5	US-10-732-923-22588
17	92.5	13.3	2468	4	US-10-755-889-615
18	92.5	13.3	2468	5	US-10-489-740-216
19	92.5	13.3	2519	5	US-10-450-763-46995
20	92	13.2	1005	4	US-10-437-963-187665
21	91	13.1	471	5	US-10-739-930-6262
22	89.5	12.9	511	4	US-10-289-762-509
23	88.5	12.7	540	5	US-10-732-923-22820
24	88.5	12.7	903	4	US-10-282-122A-52328
25	88	12.6	225	4	US-10-032-585-7829
26	88	12.6	778	4	US-10-724-972A-5663
27	88	12.6	815	5	US-10-496-905-23

28	88	12.6	1980	5	US-10-482-834A-144	Sequence 144, Appl
29	87.5	12.6	1196	4	US-10-282-122A-52737	Sequence 52737, A
30	87.5	12.6	1702	3	US-09-839-996-5	Sequence 5, Appli
31	87.5	12.6	1702	4	US-10-080-505-5	Sequence 5, Appli
32	87.5	12.6	1702	4	US-10-645-655-5	Sequence 5, Appli
33	87.5	12.6	1702	4	US-10-687-046-5	Sequence 5, Appli
34	87.5	12.6	3124	5	US-10-732-923-22709	Sequence 22709, A
35	87	12.5	281	6	US-11-097-143-21276	Sequence 21276, A
36	87	12.5	593	4	US-10-424-599-174814	Sequence 174814, A
37	87	12.5	932	4	US-10-282-122A-52510	Sequence 52510, A
38	86.5	12.4	1373	5	US-10-732-923-16976	Sequence 16976, A
39	86	12.4	145	4	US-10-424-599-149572	Sequence 149572, A
40	86	12.4	835	4	US-10-425-115-238086	Sequence 238086, A
41	86	12.4	1075	4	US-10-437-963-187664	Sequence 187664, A
42	86	12.4	1726	5	US-10-942-522-6	Sequence 6, Appli
43	86	12.4	1980	4	US-10-474-778-6	Sequence 150, Appl
44	86	12.4	1980	5	US-10-482-834A-150	Sequence 8, Appli
45	86	12.4	1980	5	US-10-942-522-8	

## ALIGNMENTS

### RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 696;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 3.6e-54;		
Matches 134;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	KEMSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKQDAGYVNLKDTFIKPVFKK	60	
Db	640	KEMSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKQDAGYVNLKDTFIKPVFKK	699	
Qy	61	IEEKKEBENKPTFVSKKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK	120	
Db	700	IEEKKEBENKPTFVSKKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK	759	
Qy	121	NNISSKSTTNPNK	134	
Db	760	NNISSKSTTNPNK	773	

### RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 696; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 60
    |||
Db 1952 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 2011
    |||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
    |||
Db 2012 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2071
    |||

QY 121 NNISKSSTNNPNK 134
    |||
Db 2072 NNISKSSTNNPNK 2085
    |||

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 696; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 60
    |||
Db 1952 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 2011
    |||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
    |||
Db 2012 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2071
    |||

QY 121 NNISKSSTNNPNK 134
    |||
Db 2072 NNISKSSTNNPNK 2085
    |||

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 696; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 60
    |||
Db 1973 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 2032
    |||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
    |||
Db 2033 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092
    |||

QY 121 NNISKSSTNNPNK 134
    |||
Db 2093 NNISKSSTNNPNK 2106
    |||

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 696; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 60
    |||
Db 1973 KEMSTTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKKGAGYVNLKDTFTFKPVFKK 2032
    |||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
    |||
Db 2033 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092
    |||

QY 121 NNISKSSTNNPNK 134
    |||
Db 2093 NNISKSSTNNPNK 2106
    |||

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Denke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169
Query Match 99.6%; Score 693; DB 5; Length 637;
Best Local Similarity 99.3%; Pred. No. 5.3e-54; Mismatches 0; Indels 0; Gaps 0;
Matches 133; Conservative 1;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKQAGYVNLKDTFKIPVFKK 60
Db 470 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKQAGYVNLKDTFKIPVFKK 529
QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
Db 530 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 589
QY 121 NNISKSTNNPNK 134
Db 590 NNISKSTNNPNK 603
; RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US2002006154A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68
Query Match 88.4%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.7e-48; Mismatches 0; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 18 YKGELEKGYQFDGWEISGFEKGKQAGYVNLKDTFKIPVFKKEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGKQAGYVNLKDTFKIPVFKKEKKKEENKPTFDVSK 60
QY 78 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 134
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 117
; RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 68
```



```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      88.4%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.7e-48;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 YKLEKGYQPDGWEISFGKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSK 77
DB 1 YKLEKGYQPDGWEISFGKGDAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSK 60
QY 78 KDNQPNVHNSQLNESHKEDLQREHSHQSKSDSTKDVATVLDKNNISSKSTNNPNK 134
DB 61 KDNQPNVHNSQLNESHKEDLQREHSHQSKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match      15.4%; Score 107; DB 5; Length 188;
Best Local Similarity 23.8%; Pred. No. 0.1;
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

QY 1 KEMSTIVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKKGAG-----YVINLSKDTF 53
DB 19 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 59
QY 54 IKPVFKKIEKKEE-----ENKPTFDVSKKKNPNQVNHNSQLNESHKEDLQ 100
DB 60 SKENDVDLDE-KEEAEETEEELKEKNEETESEISEDEEEEEEEKEEENDKKKEQEK 119
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 120 EQSNENNDQKDMA-----QNLISKQNNN 145

RESULT 9
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match      14.9%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.19;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTIVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKKGAG-----YVINLSKDTF 53
DB 1 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 41
QY 54 IKPVFKKIEKKEEENKPTFDVSKKKNPNQVNHNSQLN-----ESHKEDLQ 100
DB 42 SKENDVDLDE-KEEAEETEEELKEKNEETESEISEDEEEEEEEKEEENDKKKEQEK 100
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 101 EQSNENNDQKDMA-----QNLISKQNNN 126

RESULT 10
US-10-691-672A-3
; Sequence 3, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-3

Query Match      14.9%; Score 103.5; DB 5; Length 647;
Best Local Similarity 25.2%; Pred. No. 0.93;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTIVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKKGAG-----YVINLSKDTF 53
DB 479 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 519
QY 54 IKPVFKKIEKKEEENKPTFDVSKKKNPNQVNHNSQLN-----ESHKEDLQ 100
DB 520 SKENDVDLDE-KEEAEETEEELKEKNEETESEISEDEEEEEEEKEEENDKKKEQEK 578
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 579 EQSNENNDQKDMA-----QNLISKQNNN 604

RESULT 11
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
```



```
; NAME/KEY: MISC FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (404)..(404)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-52942

Query Match      14.3%; Score 99.5; DB 4; Length 707;
Best Local Similarity 25.3%; Pred. No. 2.4;
Matches 37; Conservative 26; Mismatches 60; Indels 23; Gaps 4;

QY 8 VSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVNLSTKTFIKPVFKKIEEK--- 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 ISIEDDAEGVKEEIDSNQDIDVV---EDKDTTDKEYDSNKEDIIPENKSKKKAKL 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 65 -----KEEN-----KPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSD 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 FGFIKKNEVRQEEENLNDISFDILDKPVENNQVKSSEIEQNELKE-IKQEEPSQHIE 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 109 SPKDVTTATVLDKNNISSKSTTNPNK 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 EERSVKIEKPIPNNDLDEKVSSNNESK 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match      14.1%; Score 98; DB 3; Length 665;
Best Local Similarity 27.1%; Pred. No. 3;
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 23 EKGYPDQGWEL--SGFEGKKDAGYVNLSTKTFIKPVFKKIEKKK-----EENK 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 EKGKQ-----DISNSNAENKDD-----VKEGVKELEKKKKEKISDDHKVBEKK 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 71 PTFD-----VSKKKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVTTATVLDKNNISS 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 KSDDHKVEENKSDDHKVEENKSDDHKTEERVKKVVEEHEEBE-----DKCKEKS 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 126 KSTTNPNK 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 ENKNKDNK 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-437-963-122282
```

```
; Sequence 122282, Application US/10437563
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRNA4530_25224C.1.pap
US-10-437-963-122282

Query Match      13.4%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 12;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY 74 DVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVTTA--TVLDKNN 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 DASKKDNHQSEGNL--SHRDEDPTRKRKKQKTNATSDACAQBWVTEKNN 665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762

Query Match      13.4%; Score 93; DB 5; Length 1529;
Best Local Similarity 27.2%; Pred. No. 23;
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

QY 47 NLSKDT-----FIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQL-----NES 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 NIISDTQCIKPIKVINSEYKKEKKEKKEKK-----NEKINTIHYSESISKNSDNEQ 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 93 H-----RKEDLQREHHSQKSDSTKDVTTATVLDKNN--ISSKSTT 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 HPFLSKLRNVKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 16
US-10-732-923-22588
; Sequence 22588, Application US/10732923
```



QY 48 LSKOTFIKPVFKKIEKKEEENKPTFDVSKKONPQVNHSQLNESHKREDLQRE-----E 102  
Best Local Similarity 13.2%; Score 92; DB 4; Length 1005;  
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;  
Db 689 VKKTKVXP-----EDKKEEKPKKEVAKKEDKTPI---KKEEKPKKEEVKKEIK 740

QY 103 HSQKSDSTKDV 113  
Db 741 KKEEKPKKEV 751

RESULT 20  
US-10-437-963-187665  
; Sequence 187665, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 187665  
; LENGTH: 1005  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_84346C.1.pap  
US-10-437-963-187665

Query Match 13.2%; Score 92; DB 4; Length 1005;  
Best Local Similarity 26.4%; Pred. No. 17;  
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;  
QY 35 GPEGKDGAGVNLKSDTFIKPVFKKIEKKEEENKPTFDVSKK-KONPQV--NHSQNLNE 91  
Db 35 GSEKEMSGKNKSIKETGTGQSKELQKGSKSRKTKDKSKKONKDMTQVPTNAEEPHK 94

QY 92 SH-----RKE-----DLQREHHSQKSDSTKDVATVLDKNNISSKST 128  
Db 95 EYTTKVRKESRTDSSIEQVIGTSSIQEMETNEQVKSOTSKDMTQVPANAEGIRKEYT 154

QY 129 T 129  
Db 155 T 155

RESULT 21  
US-10-739-930-6262  
; Sequence 6262, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 6262  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270\_1.p  
US-10-739-930-6262

Query Match 13.1%; Score 91; DB 5; Length 470;  
Best Local Similarity 19.3%; Pred. No. 8.5;  
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;  
QY 18 YKGELEKGYQPDGWEISGFE-----GKKDAGYVINLSKDTPIKPVFKKIEBKKEEN 69  
Db 100 YVQDLARRIRYDE-EATGSSQAQRIDHPNQKVGITEKAFENSPIEETSHRVDONKRINN 158

QY 70 KPTFDVSKKDN-----PQVNHSQLNE-----SHRKEDLQREHS 104  
Db 159 QKNFTAAKSSSENAVSRVSGADHKRAEVMGKPMENRDQVQTESAESKSRKENVTKSEK 218

QY 105 QKSDSTKDVATVLDKNNISSKSTNNPNK 134  
Db 219 RDQSGVKTEAKOKORNKKEKEKTSINK 248

RESULT 22  
US-10-289-762-509  
; Sequence 509, Application US/10289762  
; Publication No. US20040006218A1  
; GENERAL INFORMATION:  
; APPLICANT: Grifflais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/10/289,762  
; CURRENT FILING DATE: 2003-03-27  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 509  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-10-289-762-509

Query Match 12.9%; Score 89.5; DB 4; Length 511;  
Best Local Similarity 24.5%; Pred. No. 13;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 54 IKPVFKKIEBKKEENKPTFD-----VSKKDNPQVNHSQLNESHK 95  
Db 95 VKGVFKKTPQARPEVSSPRLPSHVQHQRPLGLEGFRDRIQKRSENPEADLGKMKRSYSD 154

QY 96 EDLQREHHSQKSDSTKDVATVLDKNNISSKST 129  
Db 155 GDLDVRVGHSDNESTEDSR-----EG3EPSPSKSS 185

RESULT 23  
US-10-732-923-22820  
; Sequence 22820, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 22820  
; LENGTH: 540  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-732-923-22820

Query Match 12.7%; Score 88.5; DB 5; Length 540;  
Best Local Similarity 24.4%; Pred. No. 17;  
Matches 31; Conservative 25; Mismatches 54; Indels 17; Gaps 4;  
QY 2 EMSSTIVSEDFILPVYKGBLEKGYQPDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK- 60



```
Db 643 SGTKVLPHSKVLMATDGLTMP-DMTGWTKEVDLAFDLTKIKVSTKGNGFVTNQISKG 701
Qy 52 TPIKVPFKKIEBKKEENKPTFDVS-----KKKONPQVNHSQLNSHRKEDLQREHSOKS 107
Db 702 QLIK-----NKDIEVLSAEDTDDQEKTEDESSONKSKKKADEHDSNTS 748
Qy 108 DSTKDVATATVLDKNNISSKSTNN 131
Db 749 SSTKN-----DKSNADSKNDSDD 766

RESULT 27
US-10-496-905-23
; Sequence 23, Application US/10496905
; Publication No. US20050192215A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Malabika
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing
; APPLICANT: Xu, Chongjun
; APPLICANT: Mulero, Julio J.
; APPLICANT: Boyle, Bryan J.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HVS-BICIP/US
; CURRENT APPLICATION NUMBER: US/10/496,905
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/598,075
; PRIOR FILING DATE: 2000-08-20
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-496-905-23

Query Match 12.6%; Score 88; DB 5; Length 815;
Best Local Similarity 27.3%; Pred. No. 31;
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;

Qy 20 GELKGYQDGCWEISG--PEGKDGAGYVNLKDTFIKPVPKKIEKKEENKPTFDVSK 77
Db 196 GQKKQESFKSWEAGKHQEVSKPAVSLQQRKQDTSKLRLPPEQKKQKQESKPSQ 255
Qy 78 -KKONPQVNHSQLNSHRKED-----LQREHSQKSDSTKVATVLDKNNISSKST 128
Db 256 WKQDTPKSKAGVQEHKKQETPKLWVPVLQKEQ-DPKKQTPKSWTPSMQSEQNTIKSWT 314
Qy 129 T 129
Db 315 T 315
```

```
RESULT 28
US-10-482-834A-144
; Sequence 144, Application US/10482834A
; Publication No. US20050074764A1
; GENERAL INFORMATION:
; APPLICANT: Mulley, John Charles
; APPLICANT: Harkin, Louise Anne
; APPLICANT: Dibbens, Michelle
; APPLICANT: Wallace, Robyn
; APPLICANT: Phillips, Hilary Ann
; APPLICANT: Heron, Sara Elizabeth
; APPLICANT: Berkovic, Samuel Frank
; APPLICANT: Scheffer, Ingrid Eileen
; APPLICANT: Bionomics Limited
; TITLE OF INVENTION: MUTATIONS IN ION CHANNELS
; FILE REFERENCE: 1386/17
; CURRENT APPLICATION NUMBER: US/10/482,834A
; CURRENT FILING DATE: 2004-01-02
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 144
; LENGTH: 1980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-482-834A-144

Query Match 12.6%; Score 88; DB 5; Length 1980;
Best Local Similarity 26.2%; Pred. No. 90;
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

Qy 27 QPDGWEISGPEGKDGAGYVNLKDTFIKPVPKKIEKKEENKPTFDVSKKONPQVNH 86
Db 993 EMNLQISVIRIKKGVAWT-KLVHAFMQAHFK---QREADEVKPLDELVEKKANCIANH 1048
Qy 87 SQLNESHKEDLQREHSQKSDSTKVATVLDKNNISSKSTNNPN 133
Db 1049 TSA-DIHRNGDFQKNGNGTTSIGSSVEKYIIDEDHM---SFNNPN 1091

RESULT 29
US-10-282-122A-52737
; Sequence 52737, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```



```

; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match          12.6%; Score 87.5; DB 3; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      46  INLSKDTPTFKPVKKLEKKKEENKPTFDVSKKONPQVNHSQLNESHRKEDIQREHS- 104
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATATETPAEKSDKQPTAASTEDASQHKANTVADNSVANNSESEPKSRRRRSI 1355
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      105  --QKSDSTKDVATVLDKNNISKSTNNPNK 134
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQTSABETTAASTDDETTIADNKRKSKPNR 1387
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 31
US-10-080-505-5
; Sequence 5, Application US/10080505
; Publication NO. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCR/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996

```

```

; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-5

Query Match          12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      46   INLSKDTPIKPVFKLKEKKKEENRPTTPVSKKONPQVNHSQLNESHKEDLOREHS- 104
         ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       1296 INTGSATAITETAERKSDPKQTETAAS TEDASOHKANTVADNSVANNSESEPKRERRSI 1355
         ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      105   -QKSDSTKVDTATVLDDKNNISKSSTNNPNK 134
         ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       1356 SQPQTSABETTAASTDETTIADNSKRSPNR 1387
         ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 32
US-10-645-655-5
; Sequence 5, Application US/10645655
; Publication No. US20040063908A1
; GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
           Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
           Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/645,655
;   FILING DATE: 20-Aug-2003
; CLASSIFICATION DATA:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/296,791
;     FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Trecartin, Richard F.
;   REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1702 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5

Query Match      12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFTKPVKFKIEEKEENKPTFDVSKKNQPQNHSQLNESHKEDLQREHS- 104
Db 1296 INTGSATATETAESKDKPQTETAATEDASQHKANTVADNSVANSSSEPKSRRRSI 1355

Qy 105 --QKSDSTKDVTATVLDDKNISKSSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNRSKRSPNR 1387

RESULT 33
US-10-687-046-5
; Sequence 5, Application US/10687046
; Publication No. US20040157241A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCS/DHR
; CURRENT APPLICATION NUMBER: US/10/687,046
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-687-046-5

Query Match      12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFTKPVKFKIEEKEENKPTFDVSKKNQPQNHSQLNESHKEDLQREHS- 104
Db 1296 INTGSATATETAESKDKPQTETAATEDASQHKANTVADNSVANSSSEPKSRRRSI 1355

Qy 105 --QKSDSTKDVTATVLDDKNISKSSTNNPNK 134
```

```

; ORGANISM: DROSOPHILA
US-11-097-143-21276

Query Match      12.5%; Score 87; DB 6; Length 281;
Best Local Similarity 21.2%; Pred. No. 11;
Matches 36; Conservative 32; Mismatches 56; Indels 46; Gaps 6;

QY 3 MSTIVSEDFILPVY-----KGELEK-----GYQFGWEISGFEKKDAGYV 45
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1 MDLSVEEQDLKNYSLLDKDNKGATSKELGMVIRALGRQPNESEVSQMINEVDSDGN 60
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 46 INLSKDTRIKPVPFKIEB-KKEENKPTFDVSKKKON-----POVNHSQ 88
   ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 61 GSIAKEEFNVLRKHNDTKNKEEKLDAFRVFDKENNGYISTTELRAVFMALEKEDDE 120
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 89 LNESHRKEDLOREHSQSDDSTKVDTA-----TVLDKNN---ISSK 126
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 121 LEEMIREYLDQDNHNINPEFTNMWTTQQDLLAKNIYSLLDKDNKGATSK 170
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 36
US-10-424-599-174814
; Sequence 174814, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 174814
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_128877C.1.pep
US-10-424-599-174814

Query Match      12.5%; Score 87; DB 4; Length 593;
Best Local Similarity 33.6%; Pred. No. 26;
Matches 37; Conservative 14; Mismatches 37; Indels 22; Gaps 8;

QY 37 EKKDAGYVINLSKDTFKIPVFKJEEKKEHENKPTFVSKKONPONVNSQL---NESH 93
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 32 EAQKDRSSNLKSKE---KDRQKEAEKKRQEKRKEK-DKGSKSGNSDVSELERRAGKESE 87
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 94 RXED-----LQREHSOKS--DSKDVATVTALDKNNISSK-TTNPNK 134
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 88 RKRFDFKXSEMRRHEH-QKSGLESKG-----QNTNNAQNKNVTANNYNR 131
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 37
US-10-282-122A-52510
; Sequence 52510, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carz, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52510
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52510

Query Match      12.5%; Score 87; DB 4; Length 932;
Best Local Similarity 24.8%; Pred. No. 45;
Matches 36; Conservative 25; Mismatches 46; Indels 38; Gaps 7;

QY 5 STIVSEDFILPVYKGELEKGYQFDGWEIFSGFEKKDAGY---VINLSKDTFKIPVFKKI 61
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 753 SELENEIVFLRKIIKGADQSY-----GIEVAKLGLPSPVINRAKE---ILQHI 799
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 62 BEKKEE-----ENKTFPDVSKKKON-----POVNHSOLNESHK---EDLORE 101
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 800 EGDKREENSLIAPSKYKDYIEVSKDTSNTKNNLGSEIKHDTLTSETNTDTTIEDBSTK 859
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 102 BHSQKSDSTKDVATVTALDKNNISSK 126
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 860 EH--LSSNKQINCRIKDESIKKE 882
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 38
US-10-732-923-16976
; Sequence 16976, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 16976
; LENGTH: 1373
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-16976

Query Match      12.4%; Score 86.5; DB 5; Length 1373;
Best Local Similarity 23.3%; Pred. No. 79;
Matches 30; Conservative 22; Mismatches 44; Indels 33; Gaps 5;

```

QY 27 QPDGM-----EISGFEGK-----DAGYVINLSKDTFIKPVFK-----KIEE 63  
Db 453 EPADWGPFTKQKEVISISITKKRPVLLHYIYVYVSVILVMDKKNFYSSAFKEIYVKIRE 512  
QY 64 KKEENKPTFDVSKKNDPQVNHSQLNESH-RKEDLQREHSQKSDSTKDVATVTLVDKNN 122  
Db 513 KOEANNKTKQITSGSNNTSSLKKNNNYDSKNKYLTTTNNKENDNT-----QNN 563  
QY 123 ISSKSTTNN 131  
Db 564 INNNNNNN 572

RESULT 39  
US-10-424-599-149572  
; Sequence 149572, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 149572  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_106086C.1.pep  
US-10-424-599-149572

Query Match 12.4%; Score 86; DB 4; Length 145;  
Best Local Similarity 26.0%; Pred. No. 5.9;  
Matches 33; Conservative 24; Mismatches 42; Indels 28; Gaps 6;  
QY 33 ISGFEGKDGAGYVINLSKDTFIKPVFKIEK--KEE-----ENKPTF---DVSKK 78  
Db 18 LSGYBERRIRSYRLQEQVQAEIIMEMLNKGPREIQSIMMVLEKQATLQQAEEKN 77  
QY 79 KNDPQVNHSQLNESHKED-----LQREHSQKSDSTKDVATVTLVDKN--NIS 124  
Db 78 ASKHQVNNNAETNNSEMEBEKTLTEMLKALQEVVLEKADTESDVNAHQADNSPINLS 137  
QY 125 SKSTTNN 131  
Db 138 SASITTS 144

RESULT 40  
US-10-425-115-238086  
; Sequence 238086, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 238086  
; LENGTH: 835  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:

; NAME/KEY: unsure  
; LOCATION: (1)..(835)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_148723C.1.pep  
US-10-425-115-238086  
Query Match 12.4%; Score 86; DB 4; Length 835;  
Best Local Similarity 25.5%; Pred. No. 48;  
Matches 37; Conservative 16; Mismatches 34; Indels 58; Gaps 7;  
QY 40 KDAGYVINLSKDTF-----IKPV-FKIEEKKBE--NKPTFDVSKKKD-----NP 82  
Db 143 KSKHIVNLEQGTSTPEVQPVAFKATEEKKKEESTPSRLPIDASKLDDEEWAHQELSANP 202  
QY 83 QVNH-----QLNESHKEDLQ-----BEHSQ 105  
Db 203 QAKEGEGLTVPFQEGLLQVCDSDRDEDKGKKMGKRYKKKGGAHMGWDFDETST 262  
QY 106 KSDSTKDVATVLDKN---NISK 126  
Db 263 DSSDEDTTNAINKGLLFPNIGHK 287  
Search completed: April 24, 2006, 15:43:32  
Job time : 68.1499 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 20.8345 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEBDFILPVYKG.....ATVLDKNNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgm2\_6/ptodata/1/iaa/5 COMB.pep.\*
- 2: /cgm2\_6/ptodata/1/iaa/6 COMB.pep.\*
- 3: /cgm2\_6/ptodata/1/iaa/H COMB.pep.\*
- 4: /cgm2\_6/ptodata/1/iaa/PCUS COMB.pep.\*
- 5: /cgm2\_6/ptodata/1/iaa/RE COMB.pep.\*
- 6: /cgm2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	696	100.0	773	2	US-09-590-991-8	Sequence 8, Appli
2	693	99.6	637	2	US-09-107-433-3169	Sequence 3169, Ap
3	693	99.6	2138	2	US-09-583-110-5274	Sequence 5274, Ap
4	615	88.4	117	2	US-08-961-083-68	Sequence 68, Appl
5	615	88.4	117	2	US-09-536-784-68	Sequence 68, Appl
6	615	88.4	117	2	US-09-765-271-68	Sequence 68, Appl
7	615	88.4	117	2	US-09-765-272A-68	Sequence 68, Appl
8	101.5	14.6	347	2	US-09-248-796A-16224	Sequence 16224, A
9	95	13.6	348	2	US-09-538-092-1316	Sequence 1316, Ap
10	92.5	13.3	2468	2	US-09-976-594-726	Sequence 726, App
11	92.5	13.3	2468	2	US-09-538-092-1135	Sequence 1135, Ap
12	92.5	13.3	2522	2	US-09-949-016-10237	Sequence 10237, A
13	89.5	12.9	511	2	US-09-198-452A-509	Sequence 509, App
14	89.5	12.9	511	2	US-09-438-185A-475	Sequence 475, App
15	88	12.6	746	2	US-09-710-279-652	Sequence 652, App
16	88	12.6	778	2	US-09-134-001C-3868	Sequence 3868, Ap
17	87.5	12.6	1702	2	US-08-296-791-5	Sequence 5, Appli
18	87.5	12.6	1702	2	US-09-839-986-5	Sequence 5, Appli
19	87.5	12.6	1702	2	US-10-080-505-5	Sequence 5, Appli
20	87.5	12.6	1702	2	US-10-645-655-5	Sequence 5, Appli
21	87.5	12.6	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
22	86	12.4	243	2	US-09-248-796A-20306	Sequence 20306, A
23	86	12.4	280	2	US-09-248-796A-17646	Sequence 17646, A
24	86	12.4	1989	2	US-09-949-016-10076	Sequence 10076, A
25	85.5	12.3	109	2	US-09-248-796A-24668	Sequence 24668, A
26	85	12.2	465	2	US-09-134-001C-3856	Sequence 3856, Ap
27	85	12.2	472	2	US-09-710-279-658	Sequence 658, App

28	85	12.2	720	2	US-09-710-279-2058	Sequence 2058, Ap
29	85	12.2	728	2	US-09-134-001C-4968	Sequence 4968, Ap
30	85	12.2	1976	2	US-09-024-020B-9	Sequence 9, Appli
31	85	12.2	1976	2	US-09-425-043-9	Sequence 9, Appli
32	85	12.2	1978	2	US-09-024-020B-3	Sequence 3, Appli
33	85	12.2	1978	2	US-09-425-043-3	Sequence 3, Appli
34	85	12.2	1988	2	US-09-024-020B-4	Sequence 4, Appli
35	85	12.2	1988	2	US-09-425-043-4	Sequence 4, Appli
36	84.5	12.1	654	2	US-10-172-502-10	Sequence 10, Appl
37	84	12.1	817	2	US-09-248-796A-20276	Sequence 20276, A
38	83.5	12.0	461	2	US-09-949-016-8508	Sequence 8508, Ap
39	83.5	12.0	476	2	US-09-248-796A-15008	Sequence 15008, A
40	83	11.9	278	2	US-09-949-016-10508	Sequence 10508, A
41	82.5	11.9	243	2	US-09-248-796A-21334	Sequence 21334, A
42	82	11.8	743	2	US-08-910-925-3	Sequence 3, Appli
43	82	11.8	743	2	US-09-949-016-6261	Sequence 6261, Ap
44	82	11.8	1024	2	US-09-270-767-44973	Sequence 44973, A
45	81.5	11.7	477	2	US-09-075-375F-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	Best Local Similarity	100.0%;	Score 696;	DB 2;	Length 773;
Matches 134;	Conservative	0;	Mismatches	0;	Indels
0;	Gaps	0;			
Qy	1	KEMSTIVSEBDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK	60		
Db	640	KEMSTIVSEBDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK	699		
Qy	61	IEEKKKEENKPTFVSKKKNPQVNSQLNSHREKDLQREHSQKSDTKDVTATVLDK	120		
Db	700	IEEKKKEENKPTFVSKKKNPQVNSQLNSHREKDLQREHSQKSDTKDVTATVLDK	759		
Qy	121	NNISSKSTNNPNK	134		
Db	760	NNISSKSTNNPNK	773		
RESULT 2	US-09-107-433-3169				
	; Sequence 3169, Application US/09107433				
	; Patent No. 6800744				
	; GENERAL INFORMATION:				
	; APPLICANT: Lynn A Doucette-Stamm and David Bush				
	; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID				
	; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE				
	; THERAPEUTICS				
	; NUMBER OF SEQUENCES: 5206				
	; CORRESPONDENCE ADDRESS:				
	; ADDRESSEE: GENOME THERAPEUTICS CORPORATION				

STREET: 100 Beaver Street	
CITY: Waltham	
STATE: Massachusetts	
COUNTRY: USA	
ZIP: 02354	
COMPUTER READABLE FORM:	
MEDIUM TYPE: CD/ROM ISO9660	
COMPUTER: <Unknown>	
OPERATING SYSTEM: <Unknown>	
SOFTWARE: <Unknown>	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/09/107,433	
FILING DATE: 30-Jun-1998	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER: 60/085131	
FILING DATE: May 12, 1998	
APPLICATION NUMBER: 60/051553	
FILING DATE: July 2, 1997	
ATTORNEY/AGENT INFORMATION:	
NAME: Ariniello, Pamela Deneke	
REGISTRATION NUMBER: 40,489	
REFERENCE/DOCKET NUMBER: GTC-011	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (781)893-5007	
TELEFAX: (781)893-8277	
INFORMATION FOR SEQ ID NO: 3169:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 637 amino acids	
TYPE: amino acid	
TOPOLOGY: linear	
MOLECULE TYPE: protein	
HYPOTHETICAL: YES	
ORIGINAL SOURCE:	
ORGANISM: Streptococcus pneumoniae	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (B) LOCATION 1...637	
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:	
US-09-107-433-3169	
Query Match	99.6%; Score 693; DB 2; Length 637;
Best Local Similarity	99.3%; Pred. No. 1.1e-66;
Matches 133; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 KEMSSSTVSEEDFILPVYKGELEKGYQPDGWEISGFEKGKQDAGYVNLKDTFKPVFKK 60
Db	470 KEMSSSTVSEEDFILPVYKGELEKGYQPDGWEISGFEKGKQDAGYVNLKDTFKPVFKK 529
Qy	61 IEKKEEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQREHSGKSDSTKDVATVLDK 120
Db	530 IEKKEEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQREHSGKSDSTKDVATVLDK 589
Qy	121 NNISSKSTNNPNK 134
Db	590 NNISSKSTNNPNK 603
RESULT 3	
US-09-583-110-5274	
Sequence 5274, Application US/09583110	
Patent No. 6699703	
GENERAL INFORMATION:	
APPLICANT: Lynn Doucette-Stamm et al.	
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus	
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics	
FILE REFERENCE: PATH00-07A	
CURRENT FILING DATE: US/09/583,110	
CURRENT FILING DATE: 2000-05-26	
PRIOR APPLICATION NUMBER: US 09/107,433	
PRIOR FILING DATE: 1998-06-30	
PRIOR APPLICATION NUMBER: US 60/085,131	
PRIOR FILING DATE: 1998-05-12	
PRIOR APPLICATION NUMBER: US 60/051,553	
Query Match	88.4%; Score 615; DB 2; Length 117;
Best Local Similarity	100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	18 YKGELEKGYQPDGWEISGFEKGKQDAGYVNLKDTFKPVFKKIEEKEEENKPTFDVSK 77

```
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIIEKKKEENKPTFDVSK 60
78 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 134
Db 61 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIIEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIIEKKKEENKPTFDVSK 60

Qy 78 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 134
Db 61 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 117

RESULT 6
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68

Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIIEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIIEKKKEENKPTFDVSK 60

Qy 78 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 134
Db 61 KDNPNVNHSQLNESHKEDLQREHSHQSDSKDVTATVLDKNNISSKSTTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```



```

; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: PB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match      88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 YKLEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKVPVKIIEKKEEENKPTFDVSK 77
DB 1 YKLEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKVPVKIIEKKEEENKPTFDVSK 60

QY 78 KDNQVNHSQLNESHKEDLQREHHSOKSDSTKDVATVLDKNNISSKSTNNPNK 134
DB 61 KDNQVNHSQLNESHKEDLQREHHSOKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-09-248-796A-16224
; Sequence 16224, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match      14.6%; Score 101.5; DB 2; Length 347;
Best Local Similarity 21.4%; Pred. No. 0.0083;
Matches 31; Conservative 32; Mismatches 63; Indels 19; Gaps 4;

QY 2 EMSSTIVSEEDFILPVYKLEKGYQPDGWEISGF-----EGKKDAGYVNLKDTFIK 55
DB 202 KLSPLMIDIDTIPRIYND--KKWTVVATSSLQNTVQTDLSESEISGWEEDDLEENYRTG 259

QY 56 PVFKIEKKEEENKPTFDVSKKDNQVNHSQLN-----ESHKEDLQREHHSOKSDS 109
DB 260 PVFKTLQLRE-----WKAKEQNPKKEEENLNQKPVAKQKPKPNSTKKQKQKQKQ 314

QY 110 TKDVTATVLDKNNISSKSTNNPNK 134
DB 315 TKKITPKTSKRMLEGISTSTNIINK 339
```

```

RESULT 9
US-09-538-092-1316
; Sequence 1316, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/539,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1316
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q14093
US-09-538-092-1316
```

```

Query Match      13.6%; Score 95; DB 2; Length 348;
Best Local Similarity 30.9%; Pred. No. 0.042;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 19 KGELEKGYQPDGWEISGPEGKDGAGYVNLKDTFIKVPVKIIEKKEEENKPTF---DV 75
DB 205 ESEGEKG---GTEKDSKKGKOS-----KKGKOSAIELQAVKADEKXDDGKKDANKGDE 256

QY 76 SK--KDNQVNHSQLN-----ESHKEDLQREHHSOKSDSTKD---VTATVLDKNNI 123
DB 257 SKDAKDAKEIKKGGKKKKPSSDSDSKDDVKE---SKDQATYDAKVAKKOTEKESA 313

QY 124 SSK 126
DB 314 DSK 316
```

```

RESULT 10
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726
```

```

Query Match      13.3%; Score 92.5; DB 2; Length 2468;
Best Local Similarity 31.0%; Pred. No. 1.1;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
```



```

US-09-134-001C-3868
; Sequence 3868, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3868
; LENGTH: 778

```

```
RESULT 18
US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSATAITETAESKDPQTETAATEDASQHKANTVADNSVANNSESSEPKSRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNINSSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 19
US-10-080-505-5
; Sequence 5, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCP/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
US-10-080-505-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSATAITETAESKDPQTETAATEDASQHKANTVADNSVANNSESSEPKSRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNINSSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 19
US-10-080-505-5
; Sequence 5, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCP/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
US-10-080-505-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSATAITETAESKDPQTETAATEDASQHKANTVADNSVANNSESSEPKSRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNINSSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 20
US-10-645-655-5
; Sequence 5, Application US/10645655
; Patent No. 6815182
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/645,655
; FILING DATE: 20-Aug-2003
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSATAITETAESKDPQTETAATEDASQHKANTVADNSVANNSESSEPKSRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNINSSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387
```

Db 1296 INTGSATALTETAEKSDKPTQETAASTEDASQHKANTVADNSVANNSESSEPKSRRRRSI 1355  
QY 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134  
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

## RESULT 21

PCT-US95-10661A-5

; Sequence 5, Application PC/TUS9510661A

; GENERAL INFORMATION:

; APPLICANT: Washington University, et al.

; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Flehr, Hobbach, Test, Albritton &amp; Herbert

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10661A

; FILING DATE: 16-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/296,791

; FILING DATE: 25-AUG-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Trecartin, Richard F.

; REGISTRATION NUMBER: 31,801

; REFERENCE/DOCKET NUMBER: FP-59941/RPT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1702 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

PCT-US95-10661A-5

Query Match 12.6%; Score 87.5; DB 4; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 2.3;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTIKPVFKKIEKEEENKPTFDVSKKONPQVNHSQLNESHKEDLOREHS- 104

Db 1296 INTGSATALTETAEKSDKPTQETAASTEDASQHKANTVADNSVANNSESSEPKSRRRRSI 1355

QY 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134

Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

## RESULT 22

US-09-248-796A-20306

; Sequence 20306, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20306

; LENGTH: 243

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20306

Query Match 12.4%; Score 86; DB 2; Length 243;  
Best Local Similarity 28.0%; Pred. No. 0.25;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY 26 YQFDGWEISGEGKKDAGYVNLKCTFIKPVFKK----IEKKKEENKPTFDVSKKK-- 79

Db 98 YDDDDDEFEGFESSGAKEINLSQAIKEWKQRRLDLEIEREKLNSKKKEEIEKAKS 157

QY 80 --DNPQVNHSQLNESHKEDLOREH--SQKSDSTKDVATVLDKNN 122

Db 158 TIDDFYENYNKRDNHQKILLQEKFISKRDDFLK--RGLWDRVN 202

## RESULT 23

US-09-248-796A-17646

; Sequence 17646, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 17646

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-17646

Query Match 12.4%; Score 86; DB 2; Length 280;  
Best Local Similarity 26.1%; Pred. No. 0.3;  
Matches 37; Conservative 23; Mismatches 50; Indels 32; Gaps 8;

QY 3 MSSTIVSEEDFILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLKCTFIKPVFKKIE 62

Db 84 VASTFCSKYDPNVSFASNLGLF-----ELYSYANKK-----KNSP--PSFEHHE 127

QY 63 -EKKEENKPTFDVSKKONPQVNHSQLNESHKEDLOREHS-----OKSDSTK 111

Db 128 IHSSEENK----YLKHPQLQRHLHLHQRVPIKSHKYEGRNTIINPIQLNDNVY 183

QY 112 DVTATVLDKNNISSKSTNNPN 133

Db 184 HINPTLLSSNG-STSTTTNNEN 204

## RESULT 24

US-09-949-016-10076

; Sequence 10076, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 1999-02-12

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10076
; LENGTH: 1989
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10076

Query Match      12.4%; Score 86; DB 2; Length 1989;
Best Local Similarity 26.2%; Pred. No. 4.1;
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

Qy 27 QPDGWEISGFGKAGYVINISKDTPIKVPKIEKKBERNKPTFDVSKKQNPQVNH 86
Db 1002 ENNLQISVIRIKKGAVMT-KLKVFHAFMQAHFK--QREADEVKPLDELYEKCANCIAH 1057
Qy 87 SOLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPN 133
Db 1058 TGA-DIHRNGDFQKNGGTGTSIGSSVEKYIIDEDHM---SPINPN 1100

RESULT 25
US-09-248-796A-24668
; Sequence 24668, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 24668
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-24668

Query Match      12.3%; Score 85.5; DB 2; Length 109;
Best Local Similarity 32.9%; Pred. No. 0.096;
Matches 27; Conservative 14; Mismatches 26; Indels 15; Gaps 4;

Qy 62 EKKKEEN-----KPTFDVSK--KQNPQVNHSQLNESHRKEDLQREHS-QKSDSTK 111
Db 11 DDDEEENTKDKNGKEDIDRNKSKSDNSNANSTQAVKNKLTETKLENHSDEKSDPTK 70
Qy 112 DVTATVLDKNNISSKSTNNPN 133
Db 71 ENS-----KQGVSKENTNAN 87

RESULT 26
US-09-134-001C-3856
; Sequence 3856, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3856
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3856

Query Match      12.2%; Score 85; DB 2; Length 465;
Best Local Similarity 32.6%; Pred. No. 0.76;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQ-----REEHSQKSDS--- 109
Db 3 MEENKQNPKE--NMSNKDDNA---THLNDSHRNEDELFRNKNNAQRERRRIDNOSK 56
Qy 110 TKDVTAT-----VLDKNNISSKSTNNPNK 134
Db 57 EKDATSQSQLETKPMDKFDNHS--HNQNK 86

RESULT 27
US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-658

Query Match      12.2%; Score 85; DB 2; Length 472;
Best Local Similarity 32.6%; Pred. No. 0.77;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQ-----REEHSQKSDS--- 109
Db 3 MEENKQNPKE--NMSNKDDNA---THLNDSHRNEDELFRNKNNAQRERRRIDNOSK 56
Qy 110 TKDVTAT-----VLDKNNISSKSTNNPNK 134
Db 57 EKDATSQSQLETKPMDKFDNHS--HNQNK 86

RESULT 28
US-09-710-279-2058
; Sequence 2058, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
```

```
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-2058

Query Match      12.2%; Score 85; DB 2; Length 720;
Best Local Similarity 28.0%; Pred. No. 1.4;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 61 TEEKKEEENKPTFDYKSKD-----NPOVNHSQLNESHKEDLQREHSQKSDSTK----D 112
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 38 LEEEQIKALDKKFKASQAKDTNKTQNNHQSNNKQNSNDKQKQSKNSKPTKCKEQN 97
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 113 VTATVLDKNNISSKSTTNPNK 134
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 98 NKGKQKNNKNTKNKQKNNK 119
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 29
US-09-134-001C-4968
; Sequence 4968, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4968
; LENGTH: 728
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4968

Query Match      12.2%; Score 85; DB 2; Length 728;
Best Local Similarity 28.0%; Pred. No. 1.4;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 61 TEEKKEEENKPTFDYKSKD-----NPOVNHSQLNESHKEDLQREHSQKSDSTK----D 112
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 46 LEEEQIKALDKKFKASQAKDTNKTQNNHQSNNKQNSNDKQKQSKNSKPTKCKEQN 105
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 113 VTATVLDKNNISSKSTTNPNK 134
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 106 NKGKQKNNKNTKNKQKNNK 127
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 30
US-09-024-020B-9
; Sequence 9, Application US/09024020B
; Patent No. 6030810
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,020B
; FILING DATE: 16-FEB-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1976 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-024-020B-9

Query Match      12.2%; Score 85; DB 2; Length 1976;
Best Local Similarity 25.2%; Pred. No. 5.3;
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

QY 27 QFDGWEISGPEGKDGAGYVNLKDTFKPVFKKIEEKKEENKPTFDVSKKKNPQVNH 86
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 989 EMNLQISVIRIKGVAVT-KVKVHAPQAHFK---QREADEVKPLDELYEKKANCANH 1044
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 87 SQLNESHKEDLQREHSQKSDSTKQVTTATVLDKNNISSKSTTNPN 133
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1045 TGV-DIHRNGDPQKNGGTSGIGSSVEKYIIDEDHM---SFINNP 1087
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 31
US-09-425-043-9
; Sequence 9, Application US/09425043
; Patent No. 6335172
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```



; APPLICATION NUMBER: US/09/425,043  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,020  
; FILING DATE: 16-FEB-1998  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1976 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-425-043-9

Query Match 12.2%; Score 85; DB 2; Length 1976;

Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

QY 27 QFDGWEISGFGKDGAGVNLNKTFTKPVFKIEKKKEENKPTPDVSKKDNQPNVH 86  
DB 989 EMNNLQISVIRIKKGAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKNCIANH 1044  
QY 87 SOLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
DB 1045 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1087

RESULT 32

US-09-024-020B-3  
; Sequence 3, Application US/09024020B  
; Patent No. 6030810  
; GENERAL INFORMATION:  
; APPLICANT: DELGADO, STEPHEN G.  
; APPLICANT: DIETRICH, PAUL S.  
; APPLICANT: FISH, LINDA M.  
; APPLICANT: HERMAN, RONALD C.  
; APPLICANT: SANGAMESWARAN, LAKSHMI  
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JANET PAULINE CLARK  
; STREET: 3401 HILLVIEW AVENUE, MS A2-250  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94304-1397  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/024,020B  
; FILING DATE: 16-FEB-1998  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1978 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-024-020B-3

Query Match 12.2%; Score 85; DB 2; Length 1978;

Best Local Similarity 25.2%; Pred. No. 5.3;

Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

QY 27 QFDGWEISGFGKDGAGVNLNKTFTKPVFKIEKKKEENKPTPDVSKKDNQPNVH 86  
DB 991 EMNNLQISVIRIKKGAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKNCIANH 1046  
QY 87 SOLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
DB 1047 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1089

RESULT 33

US-09-425-043-3  
; Sequence 3, Application US/09425043  
; Patent No. 6335172  
; GENERAL INFORMATION:  
; APPLICANT: DELGADO, STEPHEN G.  
; APPLICANT: DIETRICH, PAUL S.  
; APPLICANT: FISH, LINDA M.  
; APPLICANT: HERMAN, RONALD C.  
; APPLICANT: SANGAMESWARAN, LAKSHMI  
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JANET PAULINE CLARK  
; STREET: 3401 HILLVIEW AVENUE, MS A2-250  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94304-1397  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/425,043  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,020  
; FILING DATE: 16-FEB-1998  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1978 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide

	Query Match	12.2†	Score 85;	DB 2;	Length 1988;
	Best Local Similarity	25.2‡	Pred. No. 5.3;		
	Matches	27; Conservative	25; Mismatches	47; Indels	8; Gaps
Qy	27	QFDGWEISFGEGKKDAGVWNLSDTTFIKVPFKKIEBKKEENKPTFDSVKCKDNQPWNH	86		
Db	1001	EMNNLQLSVIRIKKGVAWT-KVKYHAFMQAHFK--ORADEVKPLDELTKKANCIAHH	1056		
Qy	87	SQLNESHREKDLOREHSQSDDSTKVATVLDKNINSSKSTTNPNP	133		

	Query Match	12.2%	Score 85	DB 2	Length 1988
	Best Local Similarity	25.4%	Pred. No. 5.3		
	Matches 27	Conservative 25	Mismatches 47	Indels 8	Gaps 4
Qy	27	QFDGWEISPEGKQAGYVILNSKQTFIKVPVKIKIEKKKEEENKPTFDVSKKDNQVNH	86		
Db	1001	EMNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKKNCIANH	1056		
Qy	87	SQLNESHKREDLQREBHSQKSDSTKDVATVLDKNNISKSTNNPN	133		
Db	1057	TGV-DIHRNGDFPKNGNGTSGIGSGSVKVKIIDEDHM---SFTNNPN	1099		

RESULT 36  
US-10-172-502-10  
; Sequence 10, Application US/10172502  
; Patent No. 6841154  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .

FILE REFERENCE: P07263US01/BAS  
CURRENT APPLICATION NUMBER: US/10/172,502  
PRIOR FILING DATE: 2002-06-17  
PRIOR APPLICATION NUMBER: US 60/298,098  
PRIOR FILING DATE: 2001-06-15  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 10  
LENGTH: 654  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-172-502-10

Query Match 12.1%; Score 84.5; DB 2; Length 654;  
Best Local Similarity 28.6%; Pred. No. 1.4;  
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;  
QY 41 DAGVYINL-SKDTETKPVFKIEEKEENKPTPDV-----SKKDNQVNHSQLNESHR 94  
DB 450 DQYHVRIVDKAFTKANTDKSNKKEQDQNSAKKEATPATPSKTPSPVKEKESQKQDSQK 509  
QY 95 KEDLQ----RREHSQKSDTKDVT--ATVLDKNNISSKSTTNNPK 134  
DB 510 DDNKQLPSVEKENDASSGKDKTPTATPTKGEVSSSTT--PTK 552

RESULT 37  
US-09-248-796A-20276  
Sequence 20276, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 20276  
LENGTH: 817  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-20276

Query Match 12.1%; Score 84; DB 2; Length 817;  
Best Local Similarity 27.0%; Pred. No. 2.1;  
Matches 24; Conservative 17; Mismatches 34; Indels 14; Gaps 4;  
QY 43 GVINLSK-----DTFIKPVFKIEEKEENKPTPDVSKKDNQVNHSQLNESHRKED 97  
DB 312 GLTISQSELDKLASTVQPILDIEGKAA--KRQIDIEKQ-----KELELQQLHEK-- 362  
QY 98 LQREHSQKSDTKDVTATVLDKNNISSK 126  
DB 363 AKGEHEAKEKEKRDIBIAKLERNQNDK 391

RESULT 38  
US-09-949-016-8508  
Sequence 8508, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8508  
LENGTH: 461  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-8508

Query Match 12.0%; Score 83.5; DB 2; Length 461;  
Best Local Similarity 21.7%; Pred. No. 1.1;  
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;  
QY 40 KDAGVVNL-SKDTETKPVFKIEEKK-----EENKPTPDVSKKKNPQ 83  
DB 26 RDSGLSQEEEDTPIE--QQLEEEKLLERERRLHBEWLLREQAQEEFRIKKEKEAA 83  
QY 84 VN-----HSQLESHRKEDLQREHSQKSDTKDVTATVLD--KNNISSKSTTNNP 132  
DB 84 KMWLEQERKLKEQWKEQQREREERKQKEKEKEEAVQKMLDQENDLENSTTWNQ 143

RESULT 39  
US-09-248-796A-15008  
Sequence 15008, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15008  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15008

Query Match 12.0%; Score 83.5; DB 2; Length 476;  
Best Local Similarity 27.6%; Pred. No. 1.1;  
Matches 27; Conservative 19; Mismatches 27; Indels 25; Gaps 3;  
QY 48 LSKDTFIKPVFKIEEKEENKPTPDVSKKKNPQVNH-----SQL 89  
DB 181 LRKDMKAK---LKLKQKQKFNDELTKDLFKKOOTPEASNPFGNSNPFNFGNLPFSKK 237  
QY 90 NESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKS 127  
DB 238 PEEKEKEKEKETSKSVADVAS-----KNAPFKS 271

RESULT 40  
US-09-949-016-10508  
Sequence 10508, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 84.0872 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTVSEEDFILPVYKG.....ATVLDKNNISSKSTNNPNK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	773	4	AAB48343
2	696	100.0	2120	3	AAY81710
3	696	100.0	2140	6	ABU01020
4	696	100.0	2140	6	ABU45746
5	696	100.0	2140	8	ADM92113
6	696	100.0	2140	8	ADT50099
7	693	99.6	637	8	ADR94534
8	693	99.6	637	9	AEA58404
9	693	99.6	2138	8	ADK48759
10	615	88.4	117	2	AAW55096
11	615	88.4	117	5	ABF54590
12	615	88.4	117	7	ADC45149
13	107	15.4	188	9	ADZ79639
14	107	15.4	354	9	ADZ72253
15	103.5	14.9	169	9	ADZ79634
16	103.5	14.9	647	9	ADZ79635
17	103.5	14.9	651	8	ADOL9012
18	103.5	14.9	651	8	ADOL9010
19	101.5	14.6	564	4	ABB61977
20	99.5	14.3	707	6	ABU25018
21	98	14.1	665	3	ABAB18278
22	98	14.1	665	7	ABO23606
23	93	13.4	1791	8	ADP25441
24	92.5	13.3	2468	6	ABR64281

25	92.5	13.3	2468	7	AD662723	Ad662723 Human Pro
26	92.5	13.3	2468	7	AD662719	Ad662719 Human Pro
27	92.5	13.3	2468	7	AD662727	Ad662727 Human Pro
28	92.5	13.3	2468	7	AD662715	Ad662715 Human Pro
29	92.5	13.3	2468	8	ADL12997	AdL12997 Human ste
30	92.5	13.3	2468	8	ADN05260	AdN05260 Antipsori
31	92.5	13.3	2468	8	ADR14614	AdR14614 Human NF-
32	92.5	13.3	2519	4	ABG16636	ABG16636 Novel hum
33	92.5	13.3	2527	8	ADN04561	AdN04561 Antipsori
34	91.5	13.1	635	8	ADS93954	AdS93954 Fibrinoge
35	91.5	13.1	635	8	ADV83292	AdV83292 Streptoco
36	91.5	13.1	643	8	ADV89902	AdV89902 Streptoco
37	91.5	13.1	643	8	ADV81155	AdV81155 Streptoco
38	91	13.1	470	8	ADT56185	AdT56185 Plant pol
39	91	13.1	484	3	AG477777	AG477777 Arabidops
40	90.5	13.0	639	9	ADM88474	AdM88474 Staphyloc
41	90	12.9	4544	8	ADP25443	ADP25443 Plasmodiu
42	89.5	12.9	292	8	ADQ66058	AdQ66058 Novel hum
43	89.5	12.9	511	2	AAV35091	AAV35091 Chlamydia
44	89.5	12.9	645	9	ADW88441	AdW88441 Staphyloc
45	88.5	12.7	903	6	ABU24404	ABU24404 Protein e

## ALIGNMENTS

### RESULT 1

AAB48343  
ID AAB48343 standard; protein; 773 AA.  
XX  
AC AAB48343;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Spi30 polypeptide.  
XX  
KW Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000WO-US015925.  
XX  
PR 10-JUN-1999; 99US-0138453P.  
XX  
PA (MEDI-) MED IMMUNE INC.  
XX  
PI Adamou JE, Choi GH;  
XX  
DR WPI; 2001-112197/12.  
XX  
DR N-PSDB; AAC84742.  
XX  
PT New vaccines comprising Spi28 or Spi30 polypeptides, for treating and  
PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections.  
XX  
PS Claim 8; Page 51-54; 54pp; English.  
XX  
CC The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Sp130 polypeptide  
XX  
SQ Sequence 773 AA;  
  
Query Match 100.0%; Score 696; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 1.1e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK 60  
DB 640 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK 699  
QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHRRKEDLQREHSQKSDSTKDVATATVLDK 120  
DB 700 IEKKKEENKPTFDVSKKONPVNHSQLNESHRRKEDLQREHSQKSDSTKDVATATVLDK 759  
QY 121 NNISKSTTNPNK 134  
DB 760 NNISKSTTNPNK 773  
  
RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1999; 98GB-00016336.  
PR 19-MAR-1999; 99US-01253299.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;  
  
Query Match 100.0%; Score 696; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK 60  
DB 1953 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKK 2012  
QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHRRKEDLQREHSQKSDSTKDVATATVLDK 120  
DB 2013 IEKKKEENKPTFDVSKKONPVNHSQLNESHRRKEDLQREHSQKSDSTKDVATATVLDK 2072  
QY 121 NNISKSTTNPNK 134  
DB 2073 NNISKSTTNPNK 2086  
  
RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 696; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYVNLKDTFIKPVFKK 60  
 DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYVNLKDTFIKPVFKK 2032  
 QY 61 IEKKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092  
 QY 121 NNISKSSTTNNPNK 134  
 DB 2093 NNISKSSTTNNPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.

XX AC ABU45746;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by Prokaryotic essential gene #31273.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Streptococcus pneumoniae.  
 XX FN WO200277183-A2.  
 XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX DR N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 73670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 2140 AA;

Query Match 100.0%; Score 696; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYVNLKDTFIKPVFKK 60  
 DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYVNLKDTFIKPVFKK 2032  
 QY 61 IEKKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092  
 QY 121 NNISKSSTTNNPNK 134  
 DB 2093 NNISKSSTTNNPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;  
 XX DT 03-JUN-2004 (first entry)  
 XX DE S pneumoniae antigenic protein sequence SeqID310.  
 XX KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 XX antigenic.  
 XX OS Streptococcus pneumoniae.



PN WO2004020609-A2.  
 XX 11-MAR-2004.  
 XX PF 02-SEP-2003; 2003WO-US027401.  
 XX PR 30-AUG-2002; 2002US-0407082P.  
 XX PA (TUFT ) UNIV TUFTS.  
 XX PI Camilli A, Hava DL;  
 XX WPI; 2004-239189/22.  
 DR N-PSDB; ADM91876.  
 XX  
 PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
 PT diagnosing, treating and preventing active infections of Streptococcus  
 PT pneumoniae.  
 XX  
 PS Claim 27; SEQ ID NO 310; 123pp; English.  
 XX  
 CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
 CC acid molecules and the antigenic polypeptides encoded by them. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
 CC compositions and methods disclosed are useful for treating Streptococcus  
 CC pneumoniae infection. The present sequence is that of an S pneumoniae  
 CC protein of the invention.  
 XX  
 SQ Sequence 2140 AA;  
 Query Match 100.0%; Score 696; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 60  
 DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 2032  
 QY 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
 QY 121 NNISKSTTNPNK 134  
 DB 2093 NNISKSTTNPNK 2106  
 RESULT 6  
 ADT50099  
 ID ADT50099 standard; protein; 2140 AA.  
 XX  
 AC ADT50099;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DZ S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
 XX  
 KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
 KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
 KW sepsis; meningitis.  
 XX  
 OS Streptococcus pneumoniae TIGR4.  
 XX  
 PN WO2004092209-A2.  
 XX  
 PD 28-OCT-2004.  
 XX  
 PF 15-APR-2004; 2004WO-EP003984.  
 XX  
 PR 15-APR-2003; 2003EP-00450087.  
 XX  
 PA (INTE-) INTERCELL AG.

XX Meinke A, Nagy B, Hanner M, Dewasthaly S, Stierschneider U;  
 XX WPI; 2004-758335/74.  
 DR N-PSDB; ADT49955.  
 XX  
 PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
 PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
 PT treating S. pneumoniae infections.  
 XX  
 PS Disclosure; SEQ ID NO 177; 191pp; English.  
 XX  
 CC This invention relates to novel nucleic acids encoding hyperimmune serum  
 CC reactive antigens, or fragments derived thereof. Specifically, it refers  
 CC to antigens selected from peptides and serum reactive epitopes that can  
 CC be used in pharmaceutical compositions that exhibit antibacterial  
 CC activity. The present invention describes a composition (including the  
 CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
 CC that is useful for manufacturing a medicament such as a vaccine, which  
 CC can be used to treat or prevent bacterial infections, particularly S.  
 CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
 CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
 CC be used for isolating, purifying and/ or identifying an interaction  
 CC partner of the hyperimmune serum reactive antigen, as well as for  
 CC manufacturing a functional nucleic acid selected from aptamers and  
 CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
 CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
 CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
 CC of the invention.  
 XX  
 SQ Sequence 2140 AA;  
 Query Match 100.0%; Score 696; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 60  
 DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 2032  
 QY 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
 QY 121 NNISKSTTNPNK 134  
 DB 2093 NNISKSTTNPNK 2106  
 RESULT 7  
 ADT94534  
 ID ADT94534 standard; protein; 637 AA.  
 XX  
 AC ADT94534;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DZ Novel S. pneumoniae protein sequence, SEQ ID 3169.  
 XX  
 KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US6800744-B1.  
 XX  
 PD 05-OCT-2004.  
 XX  
 PF 30-JUN-1998; 98US-00107433.  
 XX  
 PR 02-JUL-1997; 97US-0051553P.  
 PR 12-MAY-1998; 98US-0085131P.  
 XX

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
XX  
PT New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94809, ADR94837, ADR94969, ADR95253, ADR95642, ADR95692,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
SQ Sequence 637 AA;  
  
Query Match 99.6%; Score 693; DB 8; Length 637;  
Best Local Similarity 99.3%; Pred. No. 1.8e-62;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFK 60  
Db |||||||  
QY 470 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFK 529  
Db |||||||  
QY 61 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120  
Db |||||||  
QY 530 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 589  
Db |||||||  
QY 121 NNISKSTTNPNK 134  
Db |||||||  
QY 590 NNISKSTTNPNK 603  
Db |||||||  
  
RESULT 8  
AEA58404  
ID AEA58404 standard; protein; 637 AA.  
XX  
AC AEA58404;  
XX  
DT 25-AUG-2005 (first entry)  
XX  
DB Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
XX  
KW bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
XX vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN US2005136404-A1.  
XX

PD 23-JUN-2005.  
XX  
PF 10-JUL-2003; 2003US-00617320.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
PA (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
PT New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
PS Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
SQ Sequence 637 AA;  
  
Query Match 99.6%; Score 693; DB 9; Length 637;  
Best Local Similarity 99.3%; Pred. No. 1.8e-62;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFK 60  
Db |||||||  
QY 470 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFK 529  
Db |||||||  
QY 61 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120  
Db |||||||  
QY 530 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 589  
Db |||||||  
QY 121 NNISKSTTNPNK 134  
Db |||||||  
QY 590 NNISKSTTNPNK 603  
Db |||||||  
  
RESULT 9  
ADK48759  
ID ADK48759 standard; protein; 2138 AA.

```
XX ADK48759;
AC
DT 20-MAY-2004 (first entry)
DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
DR N-PSDB; ADK46098.
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
Query Match 99.6%; Score 693; DB 8; Length 2138;
Best Local Similarity 99.3%; Pred. No. 9.2e-62;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 60
Db 1771 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKK 2030
Qy 61 IEEKKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
Db 2031 IEEKKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2090
Qy 121 NNISKSSTNNPNK 134
Db 2091 NNISKSSTNNPNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX
AC AAW55096;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0043 protein.
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019422.
PR 31-OCT-1996; 96US-0029960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
PS Claim 11; Page 62; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKKIEEKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVPKKIEEKEENKPTFDVSK 60
Qy 78 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSSTNNPNK 134
Db 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSSTNNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
AC ABP54590;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
```

```

XX PD 23-MAY-2002.
XX PF
XX PP
XX PR 22-JAN-2001; 2001US-00765272.
XX PR 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORGs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
XX Query Match 88.4%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 2e-55;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 YKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
QY 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX FN US6573082-B1.
XX PD 03-JUN-2003.
XX PF 28-MAR-2000; 2000US-00536784.
XX PR 31-OCT-1996; 96US-0029960P.

```

```

PR 30-OCT-1997; 97US-00961083.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 59pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
XX Query Match 88.4%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 2e-55;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 YKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
QY 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
RESULT 13
ADZ79639
ID ADZ79639 standard; protein; 188 AA.
XX AC ADZ79639;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;
XX KW immunotherapy; malaria; antimalarial; vaccine.
XX OS Plasmodium falciparum.
XX FN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP ) INST PASTEUR.
XX PI Druilhe P;
XX DR WPI; 2005-355821/36.
XX

```

PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

XX Disclosure: SEQ ID NO 7; 79pp; English.

XX The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
 CC MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the  
 CC Sequence Listing is not mentioned elsewhere in the specification.

XX Sequence 188 AA;

Query Match 15.4%; Score 107; DB 9; Length 188;  
 Best Local Similarity 23.8%; Pred. No. 0.012; Mismatches 43; Indels 44; Gaps 6;  
 Matches 36; Conservative 28;  
 QY 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53  
 DB 19 KEASS-----YDYIL-----GWFGGVPVPEHKKENMLSHLYVSSKDKENI 59  
 QY 54 IKPVFKIEEKKE-----ENKPTFVSKKKNPQVNHSQLSHKSHKEDLQR 100  
 DB 60 SKENDVDLDEKEEAETEEELKEEETESIESEDEEEEBEKEEBEKKKEQEK 119  
 QY 101 EHSOKSDSTKDVATVLDKNNISSKSTNN 131  
 DB 120 EQSNENNDDQKDMA-----QNLISGNQNNN 145

RESULT 14

ADZ72253  
 ID ADZ72253 standard; protein; 354 AA.  
 XX  
 AC ADZ72253;  
 XX  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.  
 XX  
 KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
 KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN EP1526178-A1.  
 XX  
 PD 27-APR-2005.  
 XX  
 PF 24-OCT-2003; 2003EP-00292673.  
 XX  
 PR 24-OCT-2003; 2003EP-00292673.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX

PI Druilhe P;

XX WPI; 2005-323987/34.  
 DR N-PSDB; ADZ72252.

XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
 PT falciparum, which encode proteins useful for preparing vaccine  
 PT compositions against malaria.

XX Disclosure: SEQ ID NO 2; 137pp; English.

XX The present invention relates to the protection against malaria. More  
 CC particularly, the invention pertains to a family of MSP-3 (merozoite  
 CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
 CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
 CC falciparum, highly conserved in P. falciparum strains, simultaneously  
 CC expressed in P. falciparum at the erythrocytic stages and encoding  
 CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
 CC their N-terminal extremity and which are located at the merozoite  
 CC surface. The characterization of this gene family enables the definition  
 CC of immunogenic and vaccine compositions against P. falciparum. The  
 CC present sequence is the P. falciparum MSP-3-1 protein.

XX Sequence 354 AA;

Query Match 15.4%; Score 107; DB 9; Length 354;  
 Best Local Similarity 23.8%; Pred. No. 0.03; Mismatches 43; Indels 44; Gaps 6;  
 Matches 36; Conservative 28;  
 QY 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53  
 DB 185 KEASS-----YDYIL-----GWFGGVPVPEHKKENMLSHLYVSSKDKENI 225  
 QY 54 IKPVFKIEEKKE-----ENKPTFVSKKKNPQVNHSQLSHKSHKEDLQR 100  
 DB 226 SKENDVDLDEKEEAETEEELKEEETESIESEDEEEEBEKEEBEKKKEQEK 285  
 QY 101 EHSOKSDSTKDVATVLDKNNISSKSTNN 131  
 DB 286 EQSNENNDDQKDMA-----QNLISGNQNNN 311

RESULT 15

ADZ79634  
 ID ADZ79634 standard; protein; 169 AA.  
 XX  
 AC ADZ79634;  
 XX  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.  
 XX  
 KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 KW immunotherapy; malaria; antimalarial; vaccine.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..169  
 FT /note= "Amino acid residues 212-380 of MSP3"  
 XX  
 PN WO2005040206-A1.  
 XX  
 PD 06-MAY-2005.  
 XX  
 PF 22-OCT-2004; 2004WO-EP012910.  
 XX  
 PR 24-OCT-2003; 2003US-00691672.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Druilhe P;  
 XX

DR WPI; 2005-355821/36.  
 XX Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.  
 XX  
 PS Claim 2; SEQ ID NO 2; 79pp; English.  
 XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No.1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No.2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3  
 CC protein (amino acid residues 212-380).  
 XX  
 SQ Sequence 169 AA;  
 Query Match 14.9%; Score 103.5; DB 9; Length 169;  
 Best Local Similarity 25.2%; Pred. No. 0.025;  
 Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
 QY 1 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGF--EGKKDAG-----YVINLSKDTTP 53  
 Db 1 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 41  
 QY 54 IKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN-----ESHRKEDIQR 100  
 Db 42 SKENDVDLDE--KEEABETEEELKEEKEETETSEISEDEEEEEKEEKEEKKKEOEK 100  
 QY 101 EEHSQKSDTKDVTATVLDKNNISSKSTTNN 131  
 Db 101 EQSNENNQKKDMEA-----QNLISKNNNN 126  
 RESULT 16  
 ADZ79635  
 ID ADZ79635 standard; protein; 647 AA.  
 XX  
 AC ADZ79635;  
 XX  
 DT 14-JUL-2005 (first entry)  
 DE P. falciparum GLURP-MSP3 fusion protein.  
 XX  
 KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;  
 KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;  
 KW vaccine.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 PN WO2005040206-A1.  
 XX  
 PD 06-MAY-2005.  
 XX  
 PF 22-OCT-2004; 2004WO-EP012910.  
 XX  
 PR 24-OCT-2003; 2003US-00691672.

XX (INSP ) INST PASTEUR.  
 XX  
 PI Druilhe P;  
 XX  
 DR WPI; 2005-355821/36.  
 DR N-FSDB; ADZ79636.  
 XX  
 XX Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.  
 XX  
 PS Disclosure; SEQ ID NO 3; 79pp; English.  
 XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No.1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No.2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum GLURP(27-  
 CC 500)-MSP3(212-380) fusion protein.  
 XX  
 SQ Sequence 647 AA;  
 Query Match 14.9%; Score 103.5; DB 9; Length 647;  
 Best Local Similarity 25.2%; Pred. No. 0.15;  
 Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
 QY 1 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGF--EGKKDAG-----YVINLSKDTTP 53  
 Db 479 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 519  
 QY 54 IKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN-----ESHRKEDIQR 100  
 Db 520 SKENDVDLDE--KEEABETEEELKEEKEETETSEISEDEEEEEKEEKEEKKKEOEK 578  
 QY 101 EEHSQKSDTKDVTATVLDKNNISSKSTTNN 131  
 Db 579 EQSNENNQKKDMEA-----QNLISKNNNN 604  
 RESULT 17  
 ADO19012  
 ID ADO19012 standard; protein; 651 AA.  
 XX  
 AC ADO19012;  
 XX  
 DT 12-AUG-2004 (first entry)  
 DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.  
 XX  
 KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
 KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
 KW antimalarial; immunostimulant.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 PN WO2004043488-A1.





DR N-PSDB; ABL06080.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 564 AA;  
Query Match 14.6%; Score 101.5; DB 4; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.21;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 11 EDFILPVYKGELEKGYQFDGW-----EISGPEGKKDAGYVI-----NLSKDTPIK 55  
DB 78 EDLDTPLSESRSFK--VFDGWVDEHRDEHDGHDVQPSGEALDDHDDHDDHDEDEDE 135  
QY 56 PVPKTIKKEENKPT-----FDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDS 109  
DB 136 PLTELELEEEERSEPTDEDEPADEEVEEDEENNA--GENITADEABEEBEDND 193  
QY 110 TKDVTATVLDKNNISKST 128  
DB 194 EGTVEATVEATTEAT 212  
RESULT 20  
ABU25018  
ID ABU25018 standard; protein; 707 AA.  
XX  
AC ABU25018;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #10545.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Clostridium difficile.  
XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA28888.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
PS Claim 25; SEQ ID NO 52942; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 707 AA;  
Query Match 14.3%; Score 99.5; DB 6; Length 707;  
Best Local Similarity 25.3%; Pred. No. 0.45;  
Matches 37; Conservative 26; Mismatches 60; Indels 23; Gaps 4;  
QY 8 VSEEDFILPVYKGELEKGYQFDGWIEISGPEGKKDAGYVINLSKDTFIKVPKIEEK--- 64  
DB 484 ISIEDDAEEGVKEIDSNQDGVV---EDKDTTKEYDSNKEDIIEPENKSKKKAKL 540  
QY 65 -----KEEEN-----KPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSD 108  
DB 541 FGPTKKNERVEQEENLNDISPDILDKPVENNQVKSSEIEQNELKE-IKQEPESOHIE 599  
QY 109 STKDVTATVLDKNNISKSTTNNPK 134  
DB 600 EERSVKIEKPINNLDKVVSSNESK 625  
RESULT 21  
AAB18278  
ID AAB18278 standard; protein; 665 AA.  
XX  
AC AAB18278;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:135.  
XX  
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX  
OS Plasmodium falciparum.  
XX  
FN WO200025728-A2.  
XX

PD	11-MAY-2000.	
XX		
PF	05-NOV-1999; 99WO-US026796.	
XX		
PR	05-NOV-1999; 98US-0107131P.	
XX		
PA	(HOFF/) HOFFMAN S.	
PA	(CARU/) CARUCCI D.	
PA	(GARD/) GARDNER M.	
PA	(VENT/) VENTER J C.	
XX		
PI	Hoffman S, Carucci D, Gardner M, Venter JC;	
XX		
XX	WPI; 2000-365347/31.	
XX		
PT	Proteins encoded by chromosome 2 of the human malarial parasite,	
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the	
PT	diagnosis of P.falciparum infection.	
XX		
PS	Disclosure; Page 321-322; 577pp; English.	
XX		
CC	The present invention describes proteins and their fragments (I) encoded	
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.	
CC	Also described are: (i) nucleotide sequences (II) encoding (I); and (2)	
CC	vaccines against P. falciparum infection comprising (I) or (II). (I) and	
CC	(II) are useful for the development of vaccines against P. falciparum	
CC	infection. (I) and polyclonal antisera or a monoclonal antibody raised to	
CC	immunogens comprising the sequences of (I), are useful in the detection	
CC	of infection with P. falciparum. Furthermore, (I) (especially when they	
CC	are refined or secreted or membrane proteins) can aid the identification	
CC	of drugs to treat or prevent P. falciparum infection, or they can be used	
CC	to identify drug resistance in P. falciparum. Sequencing of the	
CC	Plasmodium chromosome 2 and the subsequent identification of proteins	
CC	encoded by it will help to expand our understanding of parasite biology,	
CC	a process hampered by the complexity of the parasitic lifecycle, and	
CC	provide new targets for vaccine and drug development. Parasite resistance	
CC	to drugs and mosquito resistance to insecticides have led to a resurgence	
CC	of malaria in many parts of the world, and there is a pressing need for	
CC	vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352	
CC	represent nucleotide and protein sequences given in the present	
CC	invention, but which are not specifically mentioned within the	
CC	specification	
XX		
SQ	Sequence 665 AA;	
	Query Match 14.1%; Score 98; DB 3; Length 665;	
	Best Local Similarity 27.1%; Pred. No. 0.59;	
	Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;	
Qy	23 EKGYPDGMWEI--SGFEGKKGAGYVNLKDTFFIKPVPFKIEEKE-----BENK 70	
Db	158 EKGKQ-----DISNSAENKKD-----VKEGVKELEERKKBEKISDDHKVBEVK 201	
Qy	71 PTFD----VSKKKONPQVNHSQLNESHKEDLQR--EHSQKSDSTKQVATVLDKNNISS 125	
Db	202 KSDDHKVBEENKSDDHKVEENKSKDDHKIEEVKVEEHEDEE-----DKCKEKS 252	
Qy	126 KSTTNPNK 134	
Db	253 ENKNKDENK 261	
RESULT 22		
ABO23606		
ID	ABO23606 standard; protein; 665 AA.	
XX		
AC	ABO23606;	
XX		
DT	04-SEP-2003 (first entry)	
XX		
DE	Plasmodium falciparum outlier protein #3.	
XX		
KW	Candidate protein identification; pathogen; anti-infective;	

KW	outlier protein; virulence protein; antigen; drug target protein;	
KW	pathogenic organism; antimicrobial.	
XX		
OS	Plasmodium falciparum.	
XX		
PN	US2003039963-A1.	
XX		
PD	27-FEB-2003.	
XX		
PF	30-MAR-2001; 2001US-00820843.	
XX		
PR	30-MAR-2001; 2001US-00820843.	
XX		
PA	(BRAH/) BRAHMACHARI S K.	
PA	(RAMA/) RAMACHANDRAN S.	
PA	(NAND/) NANDI T.	
PA	(BHIM/) BHIMARAO C.	
XX		
PI	Brahmachari SK, Ramachandran S, Nandi T, Bhimmarao C;	
XX		
DR	WPI; 2003-492159/46.	
XX		
PT	Identifying candidate proteins useful as anti-infectives involves	
PT	matching outlier protein sequences with protein sequences in databases.	
XX		
PS	Example 7; Page 91-93; 117pp; English.	
XX		
CC	The present invention relates to a method for identifying candidate	
CC	proteins in pathogens useful as anti-infectives. The invention discloses	
CC	a computational method which involves the calculation of several sequence	
CC	attributes and their subsequent analysis results in the identification	
CC	of outlier proteins in different pathogens. The method is useful for the	
CC	identification of outlier proteins (e.g. virulence proteins, antigens or	
CC	proteins used as drug targets) in pathogenic organisms. The method of the	
CC	invention provides reproducible results as it does not depend on the	
CC	variable biochemical characterisation of proteins. ABO23500-ABO23617	
CC	represent outlier proteins identified from different pathogenic organisms	
XX		
SQ	Sequence 665 AA;	
	Query Match 14.1%; Score 98; DB 7; Length 665;	
	Best Local Similarity 27.1%; Pred. No. 0.59;	
	Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;	
Qy	23 EKGYPDGMWEI--SGFEGKKGAGYVNLKDTFFIKPVPFKIEEKE-----BENK 70	
Db	158 EKGKQ-----DISNSAENKKD-----VKEGVKELEERKKBEKISDDHKVBEVK 201	
Qy	71 PTFD----VSKKKONPQVNHSQLNESHKEDLQR--EHSQKSDSTKQVATVLDKNNISS 125	
Db	202 KSDDHKVBEENKSDDHKVEENKSKDDHKIEEVKVEEHEDEE-----DKCKEKS 252	
Qy	126 KSTTNPNK 134	
Db	253 ENKNKDENK 261	
RESULT 23		
ADP25441		
ID	ADP25441 standard; protein; 1791 AA.	
XX		
AC	ADP25441;	
XX		
DT	09-SEP-2004 (first entry)	
XX		
DE	Plasmodium falciparum antigen amino acid sequence SEQ ID NO:18.	
XX		
KW	Plasmodium falciparum; malaria parasite; antigen; malaria; immunogenic;	
KW	immune response; cytostatic; anti-HIV; virucide; hepatotropic;	
KW	antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis;	
XX	bacterial infection.	
XX		
OS	Plasmodium falciparum.	

PV		WO2004053086-A2.
PP		24-JUN-2004.
PR		08-DEC-2003; 2003WO-US038966.
PS		06-DEC-2002; 2002US-0431494P.
PT	(EPIM-) EPIMMUNE INC. (USNA ) US SEC OF NAVY.	
PI	Sette A, Doolan DL, Carucci DJ, Sidney J, Southwood S;	
PD	WPI; 2004-468956/44.	
PE	New isolated and/or purified Plasmodium falciparum polynucleotide sequences, useful in inducing an immune response for preventing and/or treating cancer and infectious diseases, such as AIDS, hepatitis, and bacterial infections.	
PF	Claim 22; SEQ ID NO 18; 253pp; English.	
PG	The present invention describes an isolated and/or purified Plasmodium falciparum (malaria parasite) antigen polynucleotide sequence, encoding an immunogenic peptide. Also described: (1) a primer or detection probe for hybridisation with a target sequence or the amplicon generated from a target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any of the polynucleotide sequences as described above; (2) a DNA chip comprising any of the polynucleotide sequences described above; (3) a vector comprising a promoter operably linked to any of the nucleic acid sequences described above; (4) a host cell transformed by the vector of (3) or the polynucleotide described above; (5) a composition comprising a carrier and the polynucleotide described above; (6) a method of inducing an immune response in an individual comprising the administration of the polypeptide comprising any of the amino acid sequences as encoded by the polynucleotide described above; (8) a composition comprising a carrier and the polypeptide of (7); (9) a method of detecting P. falciparum in biological samples, comprising contacting a biological sample with the isolated polynucleotide and detecting the hybridisation of the isolated polynucleotides with nucleic acids contained in the sample; (10) a method for eliciting an immune response in an individual, comprising the administration of a composition comprising the polypeptides of (7) to an individual to induce an immune response in the individual; (11) an antibody that specifically binds to the P. falciparum polypeptide of (7); and (12) detecting P. falciparum antigens, comprising contacting a sample from a subject with the polypeptide of (7) and detecting the presence of an antigen-antibody complex or detecting the stimulation of T-cells in the sample. The P. falciparum antigens and immunogenic peptides have cytostatic, anti-HIV, virucide, hepatotropic and antibacterial activities, and can be used in vaccines. The methods and compositions of the present invention are useful for inducing an immune response for the prevention and/or treatment of cancer and infectious diseases, such as AIDS, hepatitis, and bacterial infections. The present sequence represents a P. falciparum antigen amino acid sequence, which is used in the exemplification of the present invention.	
PH	Sequence 1791 AA;	
PI	Query Match            13.4%; Score 93; DB 8; Length 1791; Best Local Similarity   24.7%; Pred. No. 7.5; Matches         45; Conservative   29; Mismatches   58; Indels   50; Gaps   7	
PJ	Qy      1   KEMSTTVSEBDFILPVYKGELEKYQDGFWEISG---FEGKKDAGVYNLSKDTPFKVP   57	
PK	Dd       664   KKXIKTIVSDDMFTSPVNLIKYYNVEQRKEKIVEGNLSYDKTKIPFPKFTEGRKK--   722	
PL	Qy       58   FKXIEKCEENK-----PTF-----DVSKKNQPQV   84	
PM	Dd       723   -KKIEXEKCKEKKNNNNFLVNDYSYSSPYGDENNPNFYVIYIERKDQPKCFDHPNF   781	

Qy	85	NHSQL-----NESHRK---EDLQRREHSQKSDSTKO-VTATVLDKNNISKSSTNNP	133
Dd	782	NFSKFLHNTYPMKNKKNNKNNKVRNEYNYPTSSSKDGVSYNFLSLSLFSSDNDEYSDD	841
Qy	133	NK 134	
Dd	842	NE 843	
RESULT 24			
ID	ABR64281		
ID	ABR64281	standard; protein; 2468 AA.	
XX	AC	ABR64281;	
XX	AC	XX	
DT	15-OCT-2003	(first entry)	
XX	XX		
DE	Angiogenesis protein BNO382.		
XX	XX		
KW	Cytostatic; antithematic; antiarthritic; antidiabetic; ophthalmological;		
KW	antipsoriatic; antiarteriosclerotic; cardiast; vasotropic; angiogenesis;		
KW	gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;		
KW	diabetic retinopathy; cardiovascular disease; atherosclerosis;		
KW	ischemic limb disease; coronary artery disease.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO2003027285-A1.		
XX	XX		
PD	03-APR-2003.		
XX	XX		
PF	19-SEP-2002; 2002WO-AU001282.		
XX	XX		
PR	27-SEP-2001; 2001AU-00007973.		
PR	27-SEP-2001; 2001AU-00007974.		
PR	11-OCT-2001; 2001AU-00008210.		
PR	29-OCT-2001; 2001AU-00008532.		
PR	13-NOV-2001; 2001AU-00008838.		
PR	28-AUG-2002; 2002AU-00951032.		
XX	XX		
PA	(BION-) BIONOMICS LTD.		
XX	XX		
PI	Gamble JR, Hahn CN, Vadas MA;		
XX	XX		
DR	WPI; 2003-354655/33.		
DR	N-PSDB; ACF34559.		
XX	XX		
PT	New angiogenic genes and polypeptides, useful for diagnosing,		
PT	prognosticating or treating an angiogenesis-related disorder, e.g.		
PT	cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or		
PT	cardiovascular diseases.		
XX	XX		
PS	Claim 15; SEQ ID NO 216; 90pp; English.		
XX	XX		
CC	The invention relates to the isolation of novel genes (ACF34446-ACF34559)		
CC	encoding proteins (ABR64180-ABR64281) involved in the process of		
CC	angiogenesis. The nucleic acid molecules are useful in identifying and/or		
CC	obtaining full-length human genes involved in an angiogenic process. The		
CC	nucleic acid molecule, polypeptides or complexes encoded, cells or		
CC	genetically modified non-human animals derived from these are useful for		
CC	the screening of candidate pharmaceutical compounds used in treating		
CC	angiogenesis-related disorders. They are also useful for diagnosing,		
CC	prognosticating or treating an angiogenesis-related disorder, which		
CC	involves uncontrolled or enhanced angiogenesis or is a disorder in which		
CC	a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,		
CC	diabetic retinopathy, psoriasis or cardiovascular diseases such as		
CC	atherosclerosis), or involves inappropriately arrested or decreased		
CC	angiogenesis or is a disorder in which an expanding vasculature is of		
CC	benefit (e.g. ischemic limb disease or coronary artery disease). The		
CC	modulator of expression or activity of the polypeptide encoded by the		
CC	nucleic acid sequence is useful for manufacturing a medicament for the		
CC	treatment of an angiogenesis-related disorder. This sequence corresponds		
CC	to one of the novel angiogenic protein		



subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at [ftp.wipo.int/pub/published.pct](http://wipo.int/pub/published.pct) sequences.

Query Match	13.3%	Score 92.5;	DB 7;	Length 2468;
Best Local Similarity	31.0%;	Pred. No. 13;		
Matches	22;	Conservative 19;	Mismatches 17;	Indels 13; Gaps 3;
Qy	48	LSQDTFKPVFKKIEKKEEENKPTFPVSKKQNPQVNHSQLNESHKEDLQRE-----E	102	
Db	638	VKKETVKP-----EDKKEKEKPKKEVAKKEDKTPI---KKEBKPKKEEVEKKEVKKEIK	689	
Qy	103	HSQKSDSTKDV	113	
Db	690	KBEKKBEKPKGV	700	

RESULT 27	
ADE62727	
ID	ADE62727 standard; protein; 2468 AA.
XX	
AC	ADE62727;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Human Protein AAA18904, SEQ ID NO 8660.
XX	
KW	Human; pain; neuronal tissue; gene therapy;
KW	spinal segmental nerve injury; chronic constriction injury; CCI;
KW	spared nerve injury; SNI; Chung.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at <http://wipo.int/pub/published/pct> sequences.

```
Query Match      13.3% Score 92.5; DB 7; Length 2468;  
Best Local Similarity 31.0%; Pred.No.13;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
  
Qy   48 LSKDTIKVPKKIEKKEBEENPTFDVSKCKNPQNHSQLNESHKHEDLQR-----E 102  
       :|:|||||::||::||::||::||::||::||::||::||:  
Db   638 VKKETVKP-----EKGEKEKPKEKVAKKDKTPI---KGEEKPKCEEVGKEVGKEIK 689  
  
Qy   103 HSQKS DSTKDV 113  
       ||::||  
Db   690 KEEKGPKEV 700
```

RESULT 28	
ADE62715	
ID	ADE62715 standard; protein; 2468 AA.
XX	
XX	
AC	ADE62715;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Human Protein NP_005900, SEQ ID NO 8648.
XX	
KW	Human; pain; neuronal tissue; gene therapy;
KW	spinal segmental nerve injury; chronic constriction injury; CCI;
KW	spared nerve injury; SNI; Chung.

XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
XX GENBANK; AAA18904.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX







```

PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX N-PSDB; AAS80823.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 46995; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2519 AA;
SQ
Query Match 13.3%; Score 92.5; DB 4; Length 2519;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
Oy 48 LSKDTFIKVPFKKIEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 102
Db 699 VKKETKVKP-----EDKKEEKKEPKKEVAKKEDKTPI---KKEEKPKKEEVKKEVKKEIK 740
Oy 103 HSQKSDSTKDV 113
Db 741 KEEKKEPKKEV 751
RESULT 33
ADN04561
ID ADN04561 standard; protein; 2527 AA.
XX
XX ADN04561;
XX
XX 01-JUL-2004 (first entry)
DT Antipsoriatic protein sequence #471.
DE Antipsoriatic; gene therapy; psoriasis; diagnosis.
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX Homo sapiens.
XX
XX WO2004028479-A2.
PN

```

```

XX 08-APR-2004.
XX
XX 25-SEP-2003; 2003WO-US030907.
XX
XX 25-SEP-2002; 2002US-0414006P.
XX (GETH ) GENENTECH INC.
XX
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX WPI; 2004-305105/28.
XX N-PSDB; ADN04560.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
XX Claim 9; SEQ ID NO 955; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX
XX Sequence 2527 AA;
SQ
Query Match 13.3%; Score 92.5; DB 8; Length 2527;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
Oy 48 LSKDTFIKVPFKKIEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 102
Db 697 VKKETKVKP-----EDKKEEKKEPKKEVAKKEDKTPI---KKEEKPKKEEVKKEVKKEIK 748
Oy 103 HSQKSDSTKDV 113
Db 749 KEEKKEPKKEV 759
RESULT 34
ADS93954
ID ADS93954 standard; protein; 635 AA.
XX
XX ADS93954;
XX
XX 02-DEC-2004 (first entry)
DT Fibrinogen-binding polypeptide, SEQ ID No 19.
XX
XX Fibrinogen-binding; adhesion factor; vaccine; bacterial infection;
XX Streptococcus agalactiae infection; antibacterial; gene therapy;
XX ribozyme; antisense; siRNA; anticaline; aptamer; epiegelmer.
XX
XX Streptococcus agalactiae.
XX
XX WO2004035618-A2.
XX
XX 29-APR-2004.
XX
XX 15-OCT-2003; 2003WO-EP011436.
XX
XX 15-OCT-2002; 2002EP-00023141.
PR 20-MAR-2003; 2003EP-00006393.
XX
XX (INTE-) INTERCELL AG.
XX
XX Reinscheid DJ, Gutekunst H, Schubert A, Eikmanns BJ, Meinke A;
XX WPI; 2004-357201/33.
XX

```



```

XX Streptococcus agalactiae.
OS
XX
XX FR2824074-A1.
XX
XX 31-OCT-2002.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rueniok C, Chevalier P, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 6; SEQ ID NO 2296; 2687pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
XX ADV87746-ADV89550). The nucleotide sequences encode polypeptides of S.
XX agalactiae involved in the synthesis of amino acids, cell membranes,
XX intermediate (central) metabolism, energetic metabolism, fatty acid and
XX phospholipid metabolism, nucleotide metabolism including purines,
XX pyrimidines and/or nucleosides, regulatory functions, replication,
XX transcription, translation, protein transport, adaptation to atypical
XX conditions, sensitivity to medicines and/or analogues, functions related
XX to transposons, biosynthesis of cofactors, prosthetic groups and
XX transporters, cell membrane proteins and cellular machinery. (I) are
XX useful for the detection and/or amplification of nucleic acids.
XX Pharmaceutical composition comprising (I) or (II) are useful for
XX treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
XX equivalent for the present basic patent FR2824074A1. WO200292818A2
XX contains 6617 sequence whereas the present patent only contains 2344
XX sequences.
XX
XX Sequence 643 AA;
XX
Query Match 13.1%; Score 91.5; DB 8; Length 643;
Best Local Similarity 24.5%; Pred. No. 2.7;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;
QY 3 MSSTVSEEDFTLP--VYKGEI-----EKGYQFD-----GWEISGFECKKDGAVVINLSK 50
DB 325 LTSYLENKEKFLVNPINPKNLLIREEDKYSFEDDEEFGNLLSYNKLKNEVLVPIVNTT 384
QY 51 DTFIKVPFKKIEEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKREDLQREHS--QKSD 108
DB 385 STILKP---PEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQKVEKTSPTPQKE 437
QY 109 STKQVATVLDKN--NISKSTT 129
DB 438 TVKEQTEQKVSNGTQVEKKSET 460
RESULT 37
ADV81155
ID ADV81155 standard; protein; 643 AA.
XX
XX ADV81155;
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae protein, SEQ ID 2296.
XX
XX Antibacterial; vaccine; bacterial infection.

```

---

```

XX Streptococcus agalactiae.
OS
XX
XX WO200292818-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-IB003059.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rueniok C, Chevalier P, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 6; SEQ ID NO 2296; 439pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
XX nucleotide sequences encode polypeptides of S. agalactiae involved in the
XX synthesis of amino acids, cell membranes, intermediate (central)
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,
XX regulatory functions, replication, transcription, translation, protein
XX transport, adaptation to atypical conditions, sensitivity to medicines
XX and/or analogues, functions related to transposons, biosynthesis of
XX cofactors, prosthetic groups and transporters, cell membrane proteins and
XX cellular machinery. (I) are useful for the detection and/or amplification
XX of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
XX useful for treatment of a bacterial S. agalactiae infection. The complete
XX genome of Streptococcus agalactiae is given in ADV81204. Note: The
XX present patent is an equivalent for the basic patent FR2824074A1, which
XX contains only 2344 sequences.
XX
XX Sequence 643 AA;
XX
Query Match 13.1%; Score 91.5; DB 8; Length 643;
Best Local Similarity 24.5%; Pred. No. 2.7;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;
QY 3 MSSTVSEEDFTLP--VYKGEI-----EKGYQFD-----GWEISGFECKKDGAVVINLSK 50
DB 325 LTSYLENKEKFLVNPINPKNLLIREEDKYSFEDDEEFGNLLSYNKLKNEVLVPIVNTT 384
QY 51 DTFIKVPFKKIEEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKREDLQREHS--QKSD 108
DB 385 STILKP---PEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQKVEKTSPTPQKE 437
QY 109 STKQVATVLDKN--NISKSTT 129
DB 438 TVKEQTEQKVSNGTQVEKKSET 460
RESULT 38
ADT56185
ID ADT56185 standard; protein; 470 AA.
XX
XX ADT56185;
XX
XX 13-JAN-2005 (first entry)
XX
XX Plant polypeptide, SEQ ID 6262.
XX
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;

```

KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX Viridiplantae.  
XX US2004216190-A1.  
XX 28-OCT-2004.  
XX 18-DEC-2003; 2003US-00739930.  
XX 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.  
XX (KOVA/) KOVALIC D K.  
XX Kovalic DK;  
XX WPI; 2004-757369/74.  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.

Claim 2; SEQ ID NO 6262; 14pp; English.

The invention relates a recombinant DNA construct comprising a polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO: 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean, Arabidopsis, wheat and rape but the specification does not indicate which sequences is derived from which organism. Also included is a method of producing a plant having an improved property, comprising transforming a plant with a recombinant DNA construct comprising a promoter region functional in a plant cell operably joined to a polynucleotide encoding a polypeptide associated with the property, and growing the transformed plant. The property is selected from improving plant cold tolerance, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for improving plant drought tolerance, for providing increased resistance to plant disease, for galactomannan production, for production of plant growth regulators, for improving plant heat tolerance, for improving plant tolerance to herbicides, for increasing the rate of homologous recombination in plants, for lignin production, for improving plant tolerance to extreme osmotic conditions, for improving plant tolerance to pathogens or pests, for yield improvement by modification of photosynthesis, for modifying seed oil yield and/or content, for modifying seed protein yield and/or content, for yield improvement by modification of carbohydrate, nitrogen or phosphorus use and/or uptake and for yield improvement by providing improved plant growth and development under at least one stress condition. The polynucleotide may also encode a plant transcription factor. The methods and compositions of the present invention are useful in the field of biochemistry and genetics, in particular for producing transgenic plants with improved biological characteristics such as increased yield, improved nitrogen flow, increasing plant tolerance to cold or heat, improving plant tolerance to extreme osmotic and drought conditions, and improving plant tolerance to plant pests or pathogens. They can also be used in physical arrays of molecules, plant breeding markers, computer-based storage and analysis systems. The present sequence is one of the 5544 plant protein sequences of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at [seqdata.uspto.gov/sequence.html?DocID=20040216190](http://seqdata.uspto.gov/sequence.html?DocID=20040216190).

Sequence 470 AA;

Query Match 13.1%; Score 91; DB 8; Length 470;  
Best Local Similarity 19.3%;  
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;  
OY 18 YKGELEKGYQFDGWEISGFE-----GKKDAGYVNLKSDTFIKPVFKIEKKEEN 69

Db 100 YVQDLARRIRYDE-EATGSSQAQRIDHPNCKNKGITKAFENSPIEETSHRVDNDKQINN 158  
QY 70 KPTFDVSKKDN-----PQVNHSQLNE-----SHRKEDLQREHS 104  
Db 159 QKNFTAAKSSSENAVSRVSPGADHKRAEVMGKPMENRDQVROTESAESKSHRKNVTKSEK 218  
QY 105 QKSDSTKDVTTATVLDKNNISKSTNNPNK 134  
Db 219 RDQGVKTKRAKDKRNKEKKEKTESINK 248  
RESULT 39  
AAG47777  
ID AAG47777 standard; protein; 484 AA.  
XX AAG47777;  
XX AC AAG47777;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 09-MAR-1999; 99US-0123180P.  
PR 23-MAR-1999; 99US-0123548P.  
PR 25-MAR-1999; 99US-0125788P.  
PR 29-MAR-1999; 99US-0126264P.  
PR 01-APR-1999; 99US-0126785P.  
PR 06-APR-1999; 99US-0127462P.  
PR 16-APR-1999; 99US-0128234P.  
PR 19-APR-1999; 99US-0128714P.  
PR 21-APR-1999; 99US-0129845P.  
PR 23-APR-1999; 99US-0130077P.  
PR 28-APR-1999; 99US-0130449P.  
PR 30-APR-1999; 99US-0130510P.  
PR 04-MAY-1999; 99US-0130851P.  
PR 05-MAY-1999; 99US-0131449P.  
PR 06-MAY-1999; 99US-0132048P.  
PR 11-MAY-1999; 99US-0132407P.  
PR 14-MAY-1999; 99US-0132484P.  
PR 14-MAY-1999; 99US-0132485P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132566P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 19-MAY-1999; 99US-0134768P.  
PR 20-MAY-1999; 99US-0134941P.  
PR 21-MAY-1999; 99US-0135124P.  
PR 24-MAY-1999; 99US-0135353P.  
PR 27-MAY-1999; 99US-0135629P.  
PR 27-MAY-1999; 99US-0136021P.  
PR 28-MAY-1999; 99US-0136392P.  
PR 01-JUN-1999; 99US-0136782P.  
PR 03-JUN-1999; 99US-0137222P.  
PR 04-JUN-1999; 99US-0137528P.  
PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139113P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151388P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.1%; Score 91; DB 3; Length 484;  
Best Local Similarity 19.3%; Pred. No. 2;  
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;



**THIS PAGE BLANK (user)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16, Search time 87.4631 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTTTQNGKMSSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05\_80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	748	100.0	2119	2	Q9AHT5_STRPN	Q9Aht5 streptococc
2	748	100.0	2140	2	Q97RY6_STRPN	Q97ry6 streptococc
3	745	99.6	2144	2	Q8DQP7_STRPN	Q8dqp7 streptococc
4	744	99.5	2144	2	Q9S4M8_STRPN	Q9s4m8 streptococc
5	128.5	17.2	300	2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	112	15.0	361	2	Q9SP15_PLAPA	Q9sp15 plasmodium
7	110	14.7	346	2	Q9U0G0_PLAPA	Q9u0g0 plasmodium
8	110	14.7	379	2	Q25705_PLAPA	Q25705 plasmodium
9	110	14.7	775	2	Q8CPK8_STAEP	Q8cpk8 staphylococ
10	109	14.6	3008	2	Q81436_PLAP7	Q81436 plasmodium
11	108.5	14.5	600	2	Q77355_PLAP7	Q77355 plasmodium
12	108.5	14.5	1038	2	Q90784_CHICK	Q90784 gallus gall
13	108	14.4	354	2	Q25995_PLAP7	Q25995 plasmodium
14	108	14.4	354	2	Q81J55_PLAP7	Q81j55 plasmodium
15	108	14.4	384	2	Q50VJ0_ENTHI	Q50vj0 entamoeba h
16	107	14.3	829	2	Q815F3_PLAP7	Q815f3 plasmodium
17	106.5	14.2	374	2	Q5V9M0_PLAKN	Q5v9m0 plasmodium
18	106.5	14.2	470	2	Q9FJK9_ARATH	Q9fjk9 arabidopsis
19	106	14.2	616	2	Q6BRW2_DEBHA	Q6brw2 debaromyce
20	104.5	14.0	379	2	Q9U6C4_PLAPA	Q9u6c4 plasmodium
21	104.5	14.0	380	2	Q26019_PLAPA	Q26019 plasmodium
22	104.5	14.0	736	2	Q4TVY2_PLABE	Q4tvy2 plasmodium
23	104	13.9	662	2	Q4YMU4_PLABE	Q4ymu4 plasmodium
24	104	13.9	951	2	Q96229_PLAP7	Q96229 plasmodium
25	103.5	13.8	325	2	Q44016_DICDI	Q44016 dictyosteli
26	103.5	13.8	379	2	Q25706_PLAPA	Q25706 plasmodium
27	103	13.8	296	2	Q50LX8_ENTHI	Q50lx8 entamoeba h
28	102.5	13.7	775	2	Q5HQ11_STAEO	Q5hq11 staphylococ
29	102.5	13.7	1069	2	Q512T7_ENTHI	Q512t7 entamoeba h
30	102.5	13.7	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
31	102	13.6	540	2	Q54MT2_DICDI	Q54mt2 dictyosteli

32	101.5	13.6	382	2	Q9V7J0_DROME	Q9v7j0 drosophila
33	101.5	13.6	500	2	Q6BGL7_PARTE	Q6bgl7 paramecium
34	101.5	13.6	556	2	Q9V7I9_DROME	Q9v7i9 drosophila
35	101.5	13.6	674	2	Q7RLE7_PLAYO	Q7rle7 plasmodium
36	101.5	13.6	785	2	Q9GQ82_DROME	Q9gq82 drosophila
37	101.5	13.6	954	2	Q6HNR0_BACHK	Q6hnr0 bacillus th
38	101	13.5	329	2	Q9NFV9_PLAPA	Q9nfv9 plasmodium
39	100.5	13.4	1015	2	Q5W5T1_TETPY	Q5w5t1 tetrahymena
40	100	13.4	312	2	Q9PPL5_CANJE	Q9ppl5 campylobact
41	100	13.4	400	2	Q5AL15_CANAL	Q5al15 candida alb
42	100	13.4	1011	2	Q4Y2I3_PLACH	Q4y2i3 plasmodium
43	100	13.4	1130	2	Q81JZ4_PLAP7	Q81jz4 plasmodium
44	100	13.4	1550	2	Q54GS1_DICDI	Q54gs1 dictyosteli
45	99.5	13.3	467	2	Q59PE2_CANAL	Q59pe2 candida alb

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
DOI=10.1128/JAI.69.3.1593-1598.2001;  
RA Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR001317; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

DR	TIGRFAMs; TIGR01167; LPTXG_anchor; 1.	
DR	PROSITE; PS00847; GRAM_POS_ANCHORING; 1.	
DR	PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.	
DR	PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.	
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.	
KW	Cell wall; Complete proteome; Protease.	
SQ	SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;	
Query Match 100.0%; Score 748; DB 2; Length 2140;		
Best Local Similarity 100.0%; Pred. No. 4.8e-46;		
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 HRYVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGYVNL	SK 60
Db	1963 HRYVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGYVNL	SK 2022
Qy	61 DTFIKVPVKIEKKEENKPTFDVSKKDNQVNHSQLNSHRKEDLQREHSQKSDST	120
Db	2023 DTFIKVPVKIEKKEENKPTFDVSKKDNQVNHSQLNSHRKEDLQREHSQKSDST	2082
Qy	121 KQVATVLDKNNISSKSTTNNPK	144
Db	2083 KQVATVLDKNNISSKSTTNNPK	2106
RESULT 3		
Q8DQP7_STR6	Q8DQP7_STR6 PRELIMINARY; PRT; 2144 AA.	
AC	Q8DQP7;	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).	
GN	Name-prtA; OrderedLocustNames=spr0561;	
OS	Streptococcus pneumoniae (strain AFCC BAA-255 / R6).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI_TaxID=171101;	
RN	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RX	MEDLINE=21429245; PubMed=11544234;	
RX	DOI=10.1128/JB.183.19.5709-5717.2001;	
RA	Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,	
RA	DeHoff B.S., Bateman S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,	
RA	Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,	
RA	LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,	
RA	McAhearn S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,	
RA	Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,	
RA	Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,	
RA	Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,	
RA	Glass J.I.;	
RT	"Genome of the bacterium Streptococcus pneumoniae strain R6.";	
RL	J. Bacteriol. 183:5709-5717(2001).	
RL	EMBL; AE008434; AAK9365.1; -; Genomic_DNA.	
DR	PIR; A97942; A97942.	
DR	HSSP; P00782; 2SBT.	
DR	MEROPS; S08.064; -.	
DR	GO; GO:0009986; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	GO; GO:0016020; C:membrane; IEA.	
DR	GO; GO:0008233; F:peptidase activity; IEA.	
DR	GO; GO:0042802; F:protein self binding; IEA.	
DR	GO; GO:0004289; F:subtilase activity; IEA.	
DR	GO; GO:0043086; F:negative regulation of enzyme activity; IEA.	
DR	GO; GO:0006508; F:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR010435; DUF1034.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR003137; PA.	
DR	InterPro; IPR000209; Pept_S8_S53.	
DR	InterPro; IPR010259; Prot_inh_S8a.	
DR	InterPro; IPR001680; WD40.	
DR	Pfam; PF06280; DUF1034; 1.	
DR	Pfam; PF00746; Gram_pos_anchor; 1.	

DR Pfam: PF02225; PA; 1.  
DR Pfam: PF00082; Peptidase\_S8; 1.  
DR Pfam: PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPTXG anchor; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Signal.  
FT SIGNAL 1 19 Potential.  
FT CHAIN 20 2144 cell wall-associated serine proteinase  
FT PrtA.  
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 745; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 7.9e-46;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTQNGKMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEKGDAGVYINLSK 60  
DB 1967 HRVTVTQNGKMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEKGDAGVYINLSK 2026

QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSHQKSDST 120  
DB 2027 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSHQKSDST 2086

QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 4  
Q9S4M8\_STRPN PRELIMINARY; PRT; 2144 AA.  
AC Q9S4M8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2004 (TREMBLrel. 26, Last annotation update)  
DE Cell wall-associated serine proteinase precursor PrtA.  
GN Name=prtA;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3.B;  
RX MEDLINE=21585565; PubMed=11728722;  
RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,  
RA Zysk G.;  
RT "The cell wall-associated serine proteinase PrtA: a highly conserved  
RT virulence factor of Streptococcus pneumoniae.";  
RL FEMS Microbiol. Lett. 205:99-104(2001).  
DR EMBL; AF127143; AAD48399.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SBT.  
DR MEROPS; S08.064; -.  
DR GO; GO:000986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.

DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPTXG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Signal.  
FT SIGNAL 1 19 Potential.  
FT CHAIN 20 2144 cell wall-associated serine proteinase  
FT PrtA.  
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 99.5%; Score 744; DB 2; Length 2144;  
Best Local Similarity 98.6%; Pred. No. 9.4e-46;  
Matches 142; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTQNGKMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEKGDAGVYINLSK 60  
DB 1967 HRVTVTQNGKMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEKGDAGVYINLSK 2026

QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSHQKSDST 120  
DB 2027 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSHQKSDST 2086

QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 5  
Q4XU16\_PLACH PRELIMINARY; PRT; 300 AA.  
AC Q4XU16;  
DT 13-SEP-2005 (TREMBLrel. 31, Created)  
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN ORFNames=PC000286.03.0;  
OS Plasmodium chabaudi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5825;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,  
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
RT transcriptomic, and proteomic analyses.";  
RL Science 307:82-86(2005).  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; CAJ701003049; CAH79425.1; -; Genomic\_DNA.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 4.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR ProDom; PD0000016; WD40; 3.  
DR SMART; SM00320; WD40; 5.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 4.  
DR PROSITE; PS02094; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
FT NON\_TER 1  
FT 1  
SQ SEQUENCE 300 AA; 34469 MW; 8DE4E512AFB1945 CRC64;

Query Match 17.2%; Score 128.5; DB 2; Length 300;  
Best Local Similarity 26.6%; Pred. No. 0.15;  
Matches 37; Conservative 31; Mismatches 38; Indels 33; Gaps 7;

QY 9 NGKMSSTIVSE-----EDFILPVYKGLKGYQFDGWEISGFEKGDAGVYINLSK 60

Db 140 NGKFIASSSFDKSIWISGIDGTLYAVRGHVGPAYKI-ANSI-----DNNYIVSCSQ 191  
Qy 61 DTFPIK-----PVPKLEEKKE---BENKPTFDVSKKKNQPNVNHSQLNESHKEDLQ 109  
Db 192 DSTLKLWINHLVPLLLKKKEBNAEQTKDEQK-----NEQKENPQ-NDQPNDEANSEBKK 245  
Qy 110 REEHSQKSDSTKDVATVL 128  
Db 246 KKEKEKNDKTKKIKTLL 264

## RESULT 6

ID Q95P15\_PLAFA PRELIMINARY; PRT; 361 AA.  
AC Q95P15;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FVO;  
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
RA Hisada H., Saul A., Reese J.J., Kennedy M.C., Long C.A., Miller L.H.,  
RA Stowers A.W.;  
RT "Merozoite surface protein 3 and protection against malaria in Aotus  
nancymai monkeys";  
RL J. Infect. Dis. 185:657-664 (2002).  
DR EMBL; AY044180; AAK94780.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT CHAIN <1 >361 merozoite surface protein 3.  
FT NON\_TER 1  
FT NON\_TER 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 15.0%; Score 112; DB 2; Length 361;

Best Local Similarity 23.9%; Pred. No. 3;  
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;

Qy 28 YKGELEKGYQ-----PD-----GWEISGF--EGKKDAG-----YVI 56  
Db 165 YAGKVEKDYERAKNAYQKANQAVLKAKEASSYDYLGWFGGVPFHKKENMLSHLYS 224  
Qy 57 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNQPNVNHSQLNESHKEDLQREHSQK 116  
Db 225 SKDKENISKENDVDLDE-KEEAEETEEELKEKNEETESISEDEEEEEEEKEE 283  
Qy 117 SDSTKDVATVLQKN-----NISKSTNN 141  
Db 284 NDKKKEQEKESQNNNDQKKMEAQNLISKQNNN 318

## RESULT 7

ID Q9U0G0\_PLARE PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN Name=msp3;  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;

RA Okeno D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
DR EMBL; AJ252286; CAB65754.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT NON\_TER 1  
FT NON\_TER 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPFA010 CRC64;

Query Match 14.7%; Score 110; DB 2; Length 346;

Best Local Similarity 26.2%; Pred. No. 4;  
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;

Qy 27 VYKGELEKGYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFTKPVFKKIEEKKEE 78  
Db 184 VLKAKBASSYNYILGWFGGVPFHKKENMLSHLYVSKDKENISKENDVDLDE-KEEE 242  
Qy 79 NKPTFDVSKKKNQPNVNHSQLNESHKEDLQREHSQKSDSTKDVATV-----LDKN 131  
Db 243 AETGGELEKNEETESINEDEEQEEEEEEKEENONKKEQAKESQNDQKEDMEAQ 302  
Qy 132 NISSKSTNN 141  
Db 303 NLISKQNNN 312

## RESULT 8

ID Q25705\_PLAFA PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
RA McColl D.J., Anders R.P.;  
RT "Conservation of structural motifs and antigenic diversity in the  
Plasmodium falciparum merozoite surface protein-3 (MSP-3)";  
RL Mol. Biochem. Parasitol. 90:21-31 (1997).  
DR EMBL; U08851; AAC47831.1; -; Unassigned\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 14.7%; Score 110; DB 2; Length 379;

Best Local Similarity 23.4%; Pred. No. 4.4;  
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

Qy 28 YKGELEKGYQ-----PD-----GWEISGF--EGKKDAG-----YVI 56  
Db 185 YAGKVEKDYERAKNAYQKANQAVLKAKEASSYDYLGWFGGVPFHKKENMLSHLYS 244  
Qy 57 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNQPNVNHSQLN-----ESH 103  
Db 245 SKDKENISKENDVDLDE-KEEAEETEEELKEKNEETESISEDEEEEEEEKEE 303  
Qy 104 RKEDLQREHSQKSDSTKDVATVLDKNISSKSTNN 141  
Db 304 KKEQKEQESNNNDQKKMEAE-----QNLISKQNNN 336

## RESULT 9

Q8CPK8\_STAEP PRELIMINARY; PRT; 775 AA.  
ID Q8CPK8\_STAEP  
AC Q8CPK8;

```

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocustNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL: AE016746; AAC04453.1; -; Genomic_DNA.
DR HSP; P14677; 1QME.
DR GO; GO:0008659; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP dimer.
DR InterPro; IPR012338; PBP_tpept_fold.
DR InterPro; IPR001460; Pencil_bind_tpept.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.7%; Score 110; DB 2; Length 775;
Best Local Similarity 27.0%; Pred. No. 9.5;
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

Qy 5 VTIONGKMSSTIVSEDFILPVYK-----GELEKGYQPDGW---EISGFE-----G 48
Db 625 ITIGNGKIQKQSVKSGTKVLPHSKVLMVTDGELTMP-DMTGWTKEVDLAPEDLTIKVVS 683

Qy 49 KKDAGYVNL--LSKDTTFPKVPKLEEKKEENKPTDVS-----KKKNPQVNSQLNES 102
Db 684 TKGNGFVNQSGISGQIIK-----NKKIEVLSAEDTDDQEKTDDESDSN 730

Qy 103 HRKEDLQREHSQKSDSTKVATVLDKNNISSTNN 141
Db 731 KSKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 763

RESULT 10
Q81436 PLAF7
ID Q81436 PLAF7 PRELIMINARY; PRT; 3008 AA.
AC Q81436
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein PPE0325w.
GN Name=PPE0325w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3D7;
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford T., Holroyd S., Horrocks P.,

Query Match 14.6%; Score 109; DB 2; Length 3008;
Best Local Similarity 32.4%; Pred. No. 48;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

Qy 54 YVNLK-----DTFIKVPFKIKBEKKEENKPTDVSKKONPQVNSIQ-----LNESHKKE 106
Db 2310 YDIELSKIERFGASIGVFTD--ENKKEENK--EVNKKKEENKKEENKKEENKKE 2366

Qy 107 DLQREH-----SOKSDSTKVATVLDKNNISK-----STTNPNK 144
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 11
O77355 PLAF7
ID O77355 PLAF7 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein MAL3P4.20.
GN Name=MAL3P4.20; Synonyms=PPO0465c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrall B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford T., Holroyd S., Horrocks P.,

```









```
DR Pfam: PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 379;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGF--EGKKDAG-----YVINLS 59
Db 207 VLKAKEASS-----YDYL-----GWFGGGVPEHKKEENMLSHLYVSSKD 247
QY 60 KDTFTKPVFKLKEEKKENKPTFDVSKKONPQVNHSLN-----ESHKKE 106
Db 248 KENISKENDVDLDE--KEEAEETEEELKEEKEETETSEISDEDEEEEEEKEEENKKK 306
QY 107 DLQREHSHQSKSDTKDVTATVLDDKNNISKSTNN 141
Db 307 EQEKEQSNNNDQKKDMEA-----QNLISKQNNN 336

RESULT 21
Q26019 PLAFPA PRELIMINARY; PRT; 380 AA.
AC Q26019_
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RX STRAIN=FC27;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favalaro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L07944; AAC09378.1; -; Genomic_DNA.
DR PDB; 1PSM; NMR; 0-90-127.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 380 AA; 43290 MW; 0986CA139309CA2 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 380;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGF--EGKKDAG-----YVINLS 59
Db 208 VLKAKEASS-----YDYL-----GWFGGGVPEHKKEENMLSHLYVSSKD 248
QY 60 KDTFTKPVFKLKEEKKENKPTFDVSKKONPQVNHSLN-----ESHKKE 106
Db 249 KENISKENDVDLDE--KEEAEETEEELKEEKEETETSEISDEDEEEEEEKEEENKKK 307
QY 107 DLQREHSHQSKSDTKDVTATVLDDKNNISKSTNN 141
Db 308 EQEKEQSNNNDQKKDMEA-----QNLISKQNNN 337
```

```
RESULT 22
Q4YVY2 PLABE PRELIMINARY; PRT; 736 AA.
AC Q4YVY2_
DT 13-SEP-2005 (TremBLrel. 31, Created)
DT 13-SEP-2005 (TremBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TremBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses."
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
SQ SEQUENCE 736 AA; 85816 MW; A1315CF2D97A6905 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 736;
Best Local Similarity 23.0%; Pred. No. 23;
Matches 41; Conservative 34; Mismatches 60; Indels 43; Gaps 8;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGFEGKKDAGVIVNLSDT----- 62
Db 401 INNNDSSS---GKNNAIPIFLKCKKNNKCDTIETNNPEWINKAGDVLKCKKESLTLL 457
QY 63 FIKPVFKKIEKKEE-----ENKPTFDVSKKONPQVNH-----SQLNESH----- 104
Db 458 FSSPLHKKIEKKDEKVINDEYENAKVENIKKNGRIKHPKIEFYFNEKINTPKPPDD 517
QY 105 ----KEDLQREHSHQSKSD-----STKDVAT--VLDK-----NNTSSKSTNNPNK 144
Db 518 IISANENFKTNSPFKSDILDTSVNNINANYEIEFKGDKIKNSKSVDSNIIDNDNK 575

RESULT 23
Q4YMU4 PLABE PRELIMINARY; PRT; 662 AA.
AC Q4YMU4_
DT 13-SEP-2005 (TremBLrel. 31, Created)
DT 13-SEP-2005 (TremBLrel. 31, Last sequence update)
DE RNA binding protein, putative.
GN ORFNames=PB001104.03.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses."
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
```



```
AC Q25706;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RX McColl D.J., Anders R.P.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08852; AAC47832.1; -; Unassigned_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 13.8%; Score 103.5; DB 2; Length 379;
Best Local Similarity 24.5%; Pred. No. 13;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQDGEIISGP--EGKKDAG-----YVINLS 59
DB 207 VLKAEASG-----YDYL-----GWFEFGGVPEHKKEENMLSHLYVSKD 247
QY 60 KDTFTKPKKIEKKKEENKPTFDVSKKKQNPQVNHSLN-----ESHRKE 106
DB 248 KENISKENDVDVDE-KEEAEETEEELKEEKVEETSEISDEDEEEERKEEKENDKK 306
QY 107 DLQREHSGKSDSTKDVATVLDKNNISKSTNN 141
DB 307 EQKEQSQNNENQKKDMEA-----QNLISKQNNN 336

RESULT 27
Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TReMBLrel. 31, Created)
DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1-IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Lofius B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A.,
RA Chillingworth T., Churcher K., Hance Z., Harris B., Harris D.,
RA Jagals K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Scroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-869(2005).
CC -!- CAUTION: the sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
SQ SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;
```

```
Query Match 13.8%; Score 103; DB 2; Length 296;
Best Local Similarity 28.4%; Pred. No. 11;
Matches 40; Conservative 23; Mismatches 50; Indels 28; Gaps 6;

QY 11 KEMSTIVSEEDFILPVYKGLKGYQDGEIISGFEKGKDGAGYVINLSKDTFTKPVFKK 70
DB 135 KEQKTKESG-----GDSEKKH-----DPTNEGKKN-----DTTKDNKDKKKD 177
QY 71 IEEKKE---EENKPTFDVSKKK-DNPQVNHG-----QLNESHKREKDLQREHSGKSDS 119
DB 178 TNBEGSGSGKGQKTNEESKKTENPQNNESNKEEKGKEEKKEEKGKEEKGKEE 237
QY 120 TKDVTATVLDKNNISKSTNN 140
DB 238 EQNQEKFINEKNEKENKNTN 258
DB 238 EQNQEKFINEKNEKENKNTN 258

RESULT 28
Q5HQ11 STAEQ PRELIMINARY; PRT; 775 AA.
AC Q5HQ11;
DT 10-MAY-2005 (TReMBLrel. 30, Created)
DT 10-MAY-2005 (TReMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TReMBLrel. 30, Last annotation update)
DE Penicillin-binding protein 1.
GN Name=pbpl; OrderedLocNames=SERP0746;
OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=176279;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;
RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,
RA Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J.,
RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,
RA Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,
RA Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,
RA Hance I.R., Nelson K.E., Fraser C.M.;
RT "Insights on evolution of virulence and resistance from the complete
RT genome analysis of an early methicillin-resistant Staphylococcus
RT aureus strain and a biofilm-producing methicillin-resistant
RT Staphylococcus epidermidis strain.";
RL J. Bacteriol. 187:2426-2438(2005).
DR EMBL; CP000029; AAW54126.1; -; Genomic_DNA.
DR TIGR; SERP0746; -;
DR GO; GO:0008658; P:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP_dimer.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
DR Complete proteome.
SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 775;
Best Local Similarity 24.7%; Pred. No. 34;
Matches 38; Conservative 26; Mismatches 61; Indels 29; Gaps 7;

QY 5 VTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQDGE---EISGFE-----G 48
DB 625 ITIGNGKQIKQSQSVKSGTKVLPHSKVLMVTDGELTMP-DMTGWTKEVDLAFEDLTIKVVS 683
QY 49 KKDAGYVIN--LSKDTFTKPVFK-----KIEKKKEENKPTFDVSKKKQNPQVNHQS 98
DB 684 TKGNFGFTNQSIKSGQIKKNDKITEVSLSAEDTDDQKTEDESDSKSKKDKVDENSN 743
QY 99 LNEHSHKREKDLQREHSGKSDSTKDVATVLDKNN 132
DB 744 ASSSKKNEKSNADSKNDSDDSTNETSGS--ERNN 775
```

```
RESULT 29
Q512T7 ENTHI PRELIMINARY; PRT; 1069 AA.
AC Q512T7;
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alemark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich B., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -, Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D3B8E4F0A8759CB4 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 47;
Matches 39; Conservative 31; Mismatches 54; Indels 29; Gaps 7;

QY 6 TTQNG-KEMSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKKGADGYVINLSKDTFI 64
Db 150 TLTNGEKKLSTSLCNEQD-----ELQKSSSSSTD--NKNDKRDIEHFVDVLPKNEE 199
QY 65 KVPFKKIEEKEENK-----PTFDVSKKKNQ-----VNHSLNESHRKEDLQREE 112
Db 200 KEISWEIESKTEERKSNLQIPSLNSLSEKGNKSVETAKVLKKSNSNSNGEEDKQDEE 259
QY 113 HS-QKSDSTKD-----VTATVLDKNNISSKSTT 139
Db 260 VSEKFDSEQEEKEEMIKAEVSQNKVKDKSTT 292

RESULT 30
Q9LH98 ARATH PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone: t19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones.";
```

```
RL DNA Res. 7:217-221(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002057; BAB03174.1; -, Genomic DNA.
DR HSSP; P01096; 1HF9.
DR InterPro; IPR009605; DUF1216.
DR Pfam; PF06746; DUF1216.2.
SQ SEQUENCE 2081 AA; 232852 MW; D3603E1F85EFFF29 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 2081;
Best Local Similarity 27.4%; Pred. No. 97;
Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;

QY 4 TVTQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKKGK-----AG 53
Db 1659 TVEINGGEELSTERGSKD-----GKIEBGK--EGKENSTKEGSKDKIEEGMEGKEN 1708
QY 54 YVINLSKDTFKIPVKFKIEEKEENK--TFDVS-K-KDNQPNVHSLNESHRKEDLQ- 109
Db 1709 STKESSKDGKINEIHGDKETMERGSKDGGTNTSGKSKSKSVKVEINGVDKDSLKDSKN 1769
QY 110 ---REHSQKSDSTKQVTVATVLDKNNISSKSTTNNPN 143
Db 1769 GDINEINGKEDSVKDNVTIQGNDNSLUTNSTSEPN 1805

RESULT 31
Q54WT2 DICDI PRELIMINARY; PRT; 540 AA.
AC Q54WT2;
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=DOB0186654;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugchang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Lousis H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shauleky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.P., Platzner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFI01000124; EAL64677.1; -, Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 540 AA; 60799 MW; 0F4E391CB55A35B5 CRC64;

Query Match 13.6%; Score 102; DB 2; Length 540;
Best Local Similarity 29.4%; Pred. No. 25;
Matches 25; Conservative 18; Mismatches 38; Indels 4; Gaps 1;
```

RX MEDLIN=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champagne M., Dugan S.P., Friese E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
 RA Pacleeb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA

## RESULT 33







```

RN NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- INTERACTION:
CC O46085;EG:63B12.5; NbExp=1; IntAct=EBI-123244, EBI-151469;
CC EMBL; AE003808; AAF58064.2; -; Genomic_DNA.
DR IntAct; Q9V719; -.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 556 AA; 63144 MW; B420980CBD6C357A CRC64;

Query Match 13.6%; Score 101.5; DB 2; Length 556;
Best Local Similarity 24.5%; Pred. No. 28;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 21 EDFILPVYKGELEKGYQFDGW-----EISGPEGKKDAGYVI-----NLSKDTFIK 65
DB 78 EDLDTPLSERSFK--VFDGWDEHRDHDGHDVQPSGEALDDHDEHDDHDEDEDE 135
QY 66 PVFKKIKBKEENKPT-----FDVSKKQNPQVNHSQLNESHRKDLQREHSQS KDS 119
DB 136 PLTSELELEHEERTEDEPAADBEYEEDEENNA--GNITAEADAEEREEDND 193
QY 120 TKDVTATVLDKNKISSKT 138
DB 194 EGTVEATVEATTEATT 212

RESULT 35
Q7RLE7 PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RLE7;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguioi S.V., Suh B.B., Kooij T.W., Perte M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAB101000712; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR Pfam; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PROSITE; PS00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD REPEATS 1; 2.
DR PROSITE; PS00678; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469F8E80F7 CRC64;

```

```

Query Match 13.6%; Score 101.5; DB 2; Length 674;
Best Local Similarity 23.1%; Pred. No. 34;
Matches 34; Conservative 27; Mismatches 51; Indels 35; Gaps 6;

QY 9 NGKEMSSSTIYSE-----EDFILPVYKGELEKGYQFDGWEIFSGPEGKKDAGYVINLSK 60
DB 500 NGKFIASSFDKSIIRIWSGIDGTLYLAVFRGHVGPAYXI-----DNNYIISCSQ 551
QY 61 DTFIK-----PVFKKIBE-----KKEENKPTFDV--SKKQNPQVNHSQLNE 101
DB 552 DSTLKLRINHLVPLAKKKGEGDQPKSDPKDQKNDQKDDKDDKDDKDDKNDNDHANN 611
QY 102 SHRKEDLQREHSQKSDSTKDVTATVL 128
DB 612 QEDGEEKKKKKEKDKIKSKIKITLL 638

RESULT 36
Q9GQ82 DROME PRELIMINARY; PRT; 785 AA.
ID Q9GQ82 DROME PRELIMINARY;
AC Q9GQ82;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA, isoform A).
GN Name=Asph; ORFNames=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
RT isoform of Asph missing the catalytic domain share exons with
RT junction".
RL J. Biol. Chem. 275:39543-39554(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Giannelis A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```



```

RESULT 38
QSNFV9 PLAFV
ID QSNFV9 PLAFV PRELIMINARY; PRT; 329 AA.
AC QSNFV9
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-WAR-2004 (TREMBlrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okeru D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252287; CAB85901.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF01333; Merozoite_SPAM; 1.
KW Merozoite.
FT NON TER 1
FT NON TER 329
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 13.5%; Score 101; DB 2; Length 329;
Best Local Similarity 25.3%; Pred. No. 17;
Matches 38; Conservative 25; Mismatches 47; Indels 40; Gaps 7;

Qy 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGP--EGKKDAG-----YVINLS 59
Db 171 VLKAEAS-----YDYL-----GWFGGVPHEKKERMLSHLYVSSKD 211

Qy 60 KDTFIKPVFKIEKKBEENKPTFDVSKKONQVNHSQLNESHHRKDLQREH----- 113
Db 212 KENISKENDVDVLE-KKEEABETEELEKNEERTSEISDEDEEEKEEKEQAKE 270

Qy 114 --SQKSDTKDVTATVLDKNNISSKSTNN 141
Db 271 QSNENNQKXDMEA-----QNLISKQNNN 295

RESULT 39
QSW5T1 TETPY
ID QSW5T1 TETPY PRELIMINARY; PRT; 1015 AA.
AC QSW5T1;
DT 01-FEB-2005 (TREMBlrel. 29, Created)
DT 01-FEB-2005 (TREMBlrel. 29, Last sequence update)
DT 01-FEB-2005 (TREMBlrel. 29, Last annotation update)
DE Putative P-glycoprotein (Fragment).
GN Name=mdrl;
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymenidae; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GL;
RA Camares O., Denizau F., Bamdad M.;
RT "Characterisation of MDR sequence homologue in Tetrahymena pyriformis.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AJ514918; CAD55936.2; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. ; IEA.

```

```

DR GO; GO:0000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transp_like.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00929; ABC_TM1P; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Nucleotide-binding.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1015 AA; 114219 MW; CB2E9AB73768A778 CRC64;

Query Match 13.4%; Score 100.5; DB 2; Length 1015;
Best Local Similarity 28.3%; Pred. No. 63;
Matches 41; Conservative 24; Mismatches 57; Indels 23; Gaps 6;

Qy 1 HRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVINLSK 60
Db 317 HRLT-TIENADE-----ILVIDKGLVQGTFD--QLDARGKPELAKNQIQK 362

Qy 61 DTFIKPVFKIEKKBEENKPTFDVSKKONQVNHSQLNESH-RKEDLQREHSHKSDS 119
Db 363 EQKDEERRKKEQLQEDRK--DPEERMRKSTVGKSLIMENHLTKEQIAAEKEKEQAY 419

Qy 120 TKDVTATVLDKNNISSKSTNNPNK 144
Db 420 FKE-----LDKNMVTRLFMTNRPER 439

RESULT 40
Q9PPL5 CAMJE
ID Q9PPL5 CAMJE PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative membrane protein.
GN OrderedLocusNames=Cj0692.3;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Ketley J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB72966.1; -; Genomic_DNA.
DR PIR; G81339; G81339.
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;

Query Match 13.4%; Score 100; DB 2; Length 312;
Best Local Similarity 25.3%; Pred. No. 19;
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

Qy 7 IQNGKMSSTI---VSEEDFILPVYK-----GLEKGYQFDGWISGFEKKDAGYV 55
Db 24 INQKPLDDLDLRDEISSDDILRRPFKKTPNKFLDELDDEYEYKTKSNLYLKE---L 80

Qy 56 INL----SKDTFIKPVFKIEKKBEENKPT---FDVSKKONP----QVNHSQLNESHKR 105

```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 15.1409 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTVTIQNGKEMSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: Piri:\*  
2: Piri2:\*  
3: Piri3:\*  
4: Piri4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	748	100.0	2140	2	F95074	serine proteinase,
2	745	99.6	2144	2	A97942	metalloproteinase,
3	110	14.7	558	2	T18467	hypothetical prote
4	108.5	14.5	1038	2	UJ5497	claustrin - chicke
5	104	13.9	665	2	B71609	hypothetical prote
6	103.5	13.8	325	2	T18283	hypothetical prote
7	100	13.4	312	2	G81339	probable membrane
8	99	13.2	211	2	T25911	hypothetical prote
9	97.5	13.0	1345	2	S46817	hypothetical prote
10	97	13.0	988	2	T14188	hypothetical prote
11	95.5	12.8	644	2	T47835	hypothetical prote
12	95	12.7	348	2	I37271	cylicin ii - human
13	95	12.7	535	2	T37189	hypothetical prote
14	95	12.7	622	2	A90570	lipoprotein (impor
15	95	12.7	1397	2	T10466	DNA topoisomerase
16	94.5	12.6	3724	2	T18427	hypothetical prote
17	94	12.6	210	2	T28771	hypothetical prote
18	92.5	12.4	219	2	B72291	hypothetical prote
19	92.5	12.4	540	2	D86432	hypothetical prote
20	91	12.2	253	2	T32879	hypothetical prote
21	91	12.2	385	2	T20410	hypothetical prote
22	90.5	12.1	614	2	A84152	hypothetical prote
23	90.5	12.1	670	2	T28391	ORF MSV230 hypothe
24	90.5	12.1	1332	2	S41552	probable transcrip
25	90.5	12.1	2401	2	T28676	rhoxytry protein -
26	89.5	12.0	508	2	B81594	hypothetical prote
27	89.5	12.0	508	2	E86549	hypothetical prote
28	89.5	12.0	508	2	T20744	hypothetical prote
29	89.5	12.0	700	2	S67610	probable membrane

30	89.5	12.0	1016	2	T19006	ankyrin related pr
31	89	11.9	528	2	E96795	unknown protein F2
32	89	11.9	645	2	E89883	conserved hypothet
33	89	11.9	1888	2	T39009	hypothetical prote
34	88.5	11.8	301	2	T33068	hypothetical prote
35	88.5	11.8	371	2	A71683	hypothetical prote
36	88.5	11.8	762	2	G88436	protein T04A8.13 (
37	88.5	11.8	791	2	T24435	hypothetical prote
38	88	11.8	276	2	T23451	hypothetical prote
39	88	11.8	1280	2	G96796	hypothetical prote
40	87.5	11.7	1702	2	A41859	IGA-specific metal
41	87.5	11.7	2523	2	T18477	hypothetical prote
42	87	11.6	635	2	F71621	hypothetical prote
43	87	11.6	2464	1	QKMSF1	microtubule-associ
44	87	11.6	2708	2	T09079	probable chloroqui
45	86.5	11.6	792	2	T49989	hypothetical prote

ALIGNMENTS

RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: 1-2140 <XUR>  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%;	Score	748;	DB	2;	Length	2140;
Best Local Similarity	100.0%;	Pred. No.	9.4e-49;				
Matches	144;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	HRVTVTIQNGKEMSTIVSE	DFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSK	60			
Db	1963	HRVTVTIQNGKEMSTIVSE	DFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSK	2022			
Qy	61	DTFTKPVFKIEKKKEKKEKPT	FDVSKKQNPQVNHSQLNESHKREDLQREHSQKSDST	120			
Db	2023	DTFTKPVFKIEKKKEKKEKPT	FDVSKKQNPQVNHSQLNESHKREDLQREHSQKSDST	2082			
Qy	121	KDVTATVLDKNNISSKSTNNPNK	144				
Db	2083	KDVTATVLDKNNISSKSTNNPNK	2106				

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97942

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2144 &lt;KUR&gt;

A;Cross-references: UNIPROT:Q8DQP7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;

C;Genetics:

A;Gene: prtA

C;Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 745; DB 2; Length 2144;

Best Local Similarity 99.3%; Pred. No. 1.6e-48;

Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSK 60

Db 1967 HRVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSK 2026

Qy 61 DTFIKPVFKIEEKKERENKPTFDVSKKQNPQVNHSQLNESHKREDIQRREHSQKSDST 120

Db 2027 DTFIKPVFKIEEKKERENKPTFDVSKKQNPQVNHSQLNESHKREDIQRREHSQKSDST 2086

Qy 121 KDVTATVLDKNNISSKSTNNPNK 144

Db 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 3

T18467

hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T18467

R;Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A;Reference number: Z18937

A;Accession: T18467

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-558 &lt;LAW&gt;

A;Cross-references: UNIPROT:077355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;

C;Genetics:

A;Map position: 3

A;Introns: 84/1; 160/1

A;Note: C0465c

Query Match 14.7%; Score 110; DB 2; Length 558;

Best Local Similarity 29.3%; Pred. No. 0.54;

Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

Qy 17 IVSEEDFILPVY----KGELEKGYQFDGWEISGFEGK---DAGYVINLSKDTFIKPV 67

Db 60 ILGFEDDILYECISQLKQSKRK--KADGEDKYLNAAKLKINLTGFIGNKKSDIFIEL 117

Qy 68 FKKI--BEKGER-----ENKPTDVSQ-KKQNPQVNHSQLN-----SHRK 105

Db 118 LELLNEEKGEHIAIDLNEK-TNDIKKKNENINENYENKQISNKKQKHVSHQ 176

Qy 106 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 140

Db 177 EHNINNVNLKKEKYTDIQDKRKHRSLSQSDSYKKRPFNKRKTSTIER-SLSNRYDE 235

Qy 141 NPNK 144

Db 236 KTNK 239

RESULT 4

JC5497

claustrin - chicken

N;Alternate names: keratan sulfate proteoglycan

C;Species: Gallus gallus (chicken)

C;Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 09-Jul-2004

C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.

J. Neurobiol. 25, 1-22, 1994

A;Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally

A;Reference number: JC5497; MUID:94157526; PMID:17906711

A;Accession: JC5497

A;Molecule type: mRNA

A;Residues: 1-1038 &lt;BURI&gt;

A;Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:9406318; PID:

A;Accession: PC4334

A;Molecule type: protein

A;Residues: 79-83;299-412;485-502 &lt;BUR2&gt;

A;Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5

A;Experimental source: brain

C;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo

C;Keywords: chondroin sulfate proteoglycan; glycoprotein; keratan sulfate

F;267-270/Region: cell attachment (R-G-D) motif

F;112,213,490/Binding site: carbohydrates (Asn) (covalent) #status predicted

F;152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 14.5%; Score 108.5; DB 2; Length 1038;

Best Local Similarity 28.6%; Pred. No. 1.4;

Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

Qy 11 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSKDTFIKPVK- 69

Db 595 KPETKTIVAEKDV-----TTKEEQLGKSETSEKQASEKQDVKPKVTKEKSVKCEVKA 646

Qy 70 KIEBKKEENKPTFDVSKKQNPQVNHSQLNESHKREDIQRREHSQKSDSTKDV 123

Db 647 KPEEKQDEKPKKCEVSKKEEKLPI---KKBEKPKQSDIKKEVKCKVKEEKEKKEV 702

RESULT 5

B71609

hypothetical protein PF80680w - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum

C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004

C;Accession: B71609

R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A;Reference number: A71600; MUID:99021743; PMID:9804551

A;Accession: B71609

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-665 &lt;GAR&gt;

A;Cross-references: UNIPROT:096229; UNIPARC:UPI000017B60A; GB:AE001410; GB:AE001362; NID:

A;Experimental source: clone 3D7

C;Genetics:

A;Gene: PF80680w

Query Match 13.9%; Score 104; DB 2; Length 665;

Best Local Similarity 25.9%; Pred. No. 1.9;

Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

Qy 1 HRVT-VTIQNKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFEGKDGAGYIN 57

Db 138 HRQNELNLQSGK-----NEQDI-----NKNEKGKQ-----DISNSNAENKKD----- 174

Qy 58 LSKDTFIKPVFKIEEKKK-----BENKPTFD-----VSKKQNPQVNHSQLNESH 103

Db 175 -----VKEGVKLEIEKKKEKISDD:HKVEENKSDDKHVEENKSDDKHVEENKSDDH 228

Qy 104 RKEDLQR-EHHSQKSDSTKDVTVTLVDKNNISSKSTNNPNK 144

Db 229 KIEEVKKVEHEEDEER-----DKGKKSENKKNKDNK 261

RESULT 6

T18283

hypothetical protein G5 - slime mold (Dictyostelium discoideum)

C;Species: Dictyostelium discoideum







C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: A90570  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, P.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: A90570  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-622 <KUR>  
A;Cross-references: UNIPROT:Q980A1; UNIPARC:UPI00000C80A1; GB:AL445566; PID:g14089879; EMBL:Z98547; NID:e1325376; F22291  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPU 4650  
A;Genetic code: SGC3

Query Match 12.7%; Score 95; DB 2; Length 622;  
Best Local Similarity 25.4%; Pred. No. 8.3;  
Matches 34; Conservative 26; Mismatches 50; Indels 24; Gaps 5;  
Qy 18 VSREDFILPVYKGELEKGYQFDGWEISGFEKGKQAGYVINLSKDTFIKPVFKKIEKKKE 77  
Db LSEDNILFSLNIENNAKQLDEP-----VSQEKPKIKFQBINFSQTE 128  
Qy 78 ENKPTFVDSKKD-----NPQVHSQLNES--HRKEDLQREHSHQSDSTKQVATVLDKN 131  
Db 129 Q-KITDNISSEDEKKNPKDNENSNNSNSQKDELQKNSSDKLNDNVQDEKANKENS 187  
Qy 132 -NISKSSTTNPNK 144  
Db 188 SNDSKKKQNDENTNK 201

RESULT 15  
T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C;Accession: T10466  
R;Cheesman, S.J.  
A;Title: submitted to the EMBL Data Library, September 1995  
A;Reference number: 217031  
A;Accession: T10466  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1397 <CHE>  
A;Cross-references: UNIPARC:UPI00000006D7; EMBL:X79345; NID:g594807  
C;Genetics:  
A;Gene: TopoII  
A;Map position: 14  
C;Superfamily: eukaryotic type II DNA topoisomerase; phase T4 DNA topoisomerase (ATP-hydrolyzing)  
C;Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 12.7%; Score 95; DB 2; Length 1397;  
Best Local Similarity 23.8%; Pred. No. 21;  
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;  
Qy 9 NGKEMSSSTIVSEDFIL--PVYKGELEKGYQFDGWEISGFEKGKQAGYVINLSKDTFIKP 66  
Db 1132 NEETIAGITVDYDYLSPFSLTLEK---VEDLLTQLEKERELELRNITVETWMLK 1188  
Qy 67 VPKKIE-----KKEENKPTPDVSKKQNDPQVNHSQLMESHRKEDLQREHSHQ 115  
Db 1189 DIEKVEAIEPQNVLSNREESNK--FKVARKQ-----GFSSMKKKKKKSSDSESE 1241  
Qy 116 ---KSDSTKQVATVLDKNNISSSKSTTNPN 143  
Db 1242 GGDTSDSSEFLVNTLNKKNTNKKTTTSSN 1272

RESULT 16  
T18427  
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18427  
R;Lawson, D.; Bowman, S.; Barrell, B.  
A;Title: submitted to the EMBL Data Library, August 1997  
A;Reference number: 218935  
A;Accession: T18427  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-3724 <LAW>  
A;Cross-references: UNIPROT:O77320; UNIPARC:UPI000007935D; EMBL:Z98547; NID:e1325376; F22291  
C;Genetics:  
A;Introns: 307/1; 1545/2  
A;Note: C0335c

Query Match 12.6%; Score 94.5; DB 2; Length 3724;  
Best Local Similarity 22.1%; Pred. No. 67;  
Matches 32; Conservative 30; Mismatches 50; Indels 33; Gaps 6;  
Qy 1 HRVTVTIONGKEMSSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGKQAGYVINLSK 60  
Db 1046 HKIEQDIIQIHISQTNICDNN-IEQINRENSKKGVRIISGTDN---ENKND----- 1092  
Qy 61 DTFIKPVFKKIEKKKEENKPTPDVSKKQNDPQVNH-SQLNESHRKEDLQREHSHQSDS 119  
Db 1093 -----MENKNDMEKKN--DMKKNDIEKKNDMEKKNDMEKKNDMEKKNDMEKKND 1140  
Qy 120 TKQVATVLDKNNISSSKSTTNPNK 144  
Db 1141 ME-----NENNNENKSDIENENK 1158

RESULT 17  
T28771  
hypothetical protein E03H12.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T28771  
R;Nelson, J.; Wohldmann, P.; Sansone, J.  
A;Title: submitted to the EMBL Data Library, June 1997  
A;Description: The sequence of C. elegans cosmid E03H12.  
A;Reference number: 220520  
A;Accession: T28771  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-210 <NEL>  
A;Cross-references: UNIPROT:O02124; UNIPARC:UPI000007D7F6; EMBL:AF000299; PIDN:AAC47980  
A;Experimental source: strain Bristol N2; clone E03H12  
C;Genetics:  
A;Gene: CESP:E03H12.5  
A;Map position: 4  
A;Introns: 30/2; 201/3

Query Match 12.6%; Score 94; DB 2; Length 210;  
Best Local Similarity 28.6%; Pred. No. 3;  
Matches 28; Conservative 18; Mismatches 52; Indels 0; Gaps 0;  
Qy 47 EGKKDAGYVINLSKDTFIKPVFKKIEKKKEENKPTPDVSKKQNDPQVNHSQLMESHRKE 106  
Db 69 EGEKKQGGKK 128  
Qy 107 DLQREHSHQSDSTKQVATVLDKNNISSSKSTTNPNK 144  
Db 129 EKQDKKQDKKK 166

RESULT 18  
B72291  
hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: B72291  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke





Db 549 ITRNRADLEKEI-----EDM-----EKDKOYELDEBEVAGSRGK-----LNMGAHMLA 594  
Qy 65 K--PVEKKIEEKEENKPTFD-----VSKKDD-----NPQVNHSQLNESH 103  
Db 595 KENGKVSQSSKTKVDRAPTNDKLTSVIPEGEKEKDKTASSVTTVHENVNKNKEIKENG 654  
Qy 104 RKEDLQR-EEHSQKSDSTKDVATVLD-KNNISSKSTTN 140  
Db 655 KNEEQDMVEESSKTESKDAADAAKQDYGLODKTAEN 693  
  
RESULT 25  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C:Species: Plasmodium yoelii  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28676; A45521  
R:Singha, K.A.; Keen, J.K.; Ogund, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2401 <SIN>  
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:G1041784; PI  
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KE>  
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281  
  
Query Match 12.1%; Score 90.5; DB 2; Length 2401;  
Best Local Similarity 27.4%; Pred. No. 82;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;  
  
Qy 6 TQNGKMSSTI-----VSEDFILPVKGELEKGY---QFDGWEISGFEKGKDA 52  
Db 924 TSKNHEKISDIRKNSLKIIDFSESYINDI-KKELEKNVLESQNNNTDINQVLSKIEN 982  
  
Qy 53 GVINLSKDTFTKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE 112  
Db 983 IF--NILKLNKIKKIIDKVKXYTDIEK-----NNKINAELSNSEKIITQLKENSLSKE 1035  
  
Qy 113 HSQKSDSTKD-----VTATVLDKNNISS 135  
Db 1036 QSKIKSTIDNNYVSECINKNTNLKTYIVNEKNNT 1072  
  
RESULT 26  
B81594  
hypothetical protein CP0281 [imported] - Chlamydia pneumoniae (strain AR39)  
C:Species: Chlamydia pneumoniae  
C:Accession: B81594  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: B81500; MUID:20150255; PMID:10684935  
A:Accession: B81594  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <REA>  
A:Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:AE002189; GB:AE002161; NID  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0281

Query Match 12.0%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 17;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
  
Qy 64 IKPVFKKIEEKEENKPTFD-----VSKKDNQVNHSQLNESHK 105  
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151  
  
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139  
Db 152 GDLDVRGHDSNEDSTEDSR-----EGGEPSKSSS 182  
  
RESULT 27  
B86549  
hypothetical protein CP10473 [imported] - Chlamydia pneumoniae (strain J138)  
C:Species: Chlamydia pneumoniae  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: B86549  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: B86549  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <STO>  
A:Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:BA000008; NID:G8978843; PI  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CP10473  
  
Query Match 12.0%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 17;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
  
Qy 64 IKPVFKKIEEKEENKPTFD-----VSKKDNQVNHSQLNESHK 105  
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151  
  
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139  
Db 152 GDLDVRGHDSNEDSTEDSR-----EGGEPSKSSS 182  
  
RESULT 28  
C72074  
hypothetical protein - Chlamydia pneumoniae (strain CWL029)  
C:Species: Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: C72074  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: C72074  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <ARN>  
A:Cross-references: UNIPROT:Q92878; UNIPARC:UPI00000C11CC; GB:AE001632; GB:AE001363; NID  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: CPn0473  
  
Query Match 12.0%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 17;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
  
Qy 64 IKPVFKKIEEKEENKPTFD-----VSKKDNQVNHSQLNESHK 105  
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151  
  
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139

Query Match 12.0%; Score 89.5; DB 2; Length 1016;  
Best Local Similarity 21.6%; Pred. No. 38;  
Matches 33; Conservative 33; Mismatches 38; Indels 49; Gaps 8

QY 5 VTIQNKMSSTIVSEEDFILPVYKGELEKGYQPGWEISGFGKGDAGVNLNKDTFI 64  
DB 265 ILLENGALSD-----LTFTGADVIGVADKCIDYVELA-DTV-357  
QY 65 KPVFKKIEKKKEE---NKPTFDVSKKONPOVNHQ-LNESHRKEDLQ-REEHSOK---116  
DB 303 -----KVQNKKS PGSGQPPTSIIQKNHRMPSHEEHVLTSEKEDLQHKQHSNEFL 357  
QY 117 -----SDSTKDVAT-VLDKNNISS 135  
DB 358 HSHFSTASVGSTTSNTNTTTTIVIGENDISA 390

RESULT 31  
E96795  
unknown protein F28016.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: E96795  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
Article: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E96795  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-528 <STO>  
A:Cross-references: UNIPROT:Q9SRE2; UNIPARC:UPI00000A4C99; GB:AE005173; NID:961  
C:Genetics:  
A:Gene: F28016.8  
A:Map position: 1

```

Amap position: 1
      Query Match      11.9%; Score 89; DB 2; Length 528;
      Best Local Similarity 24.3%; Pred. No. 20;
      Matches 35; Conservative 22; Mismatches 51; Indels 36; Gaps 6

Qy 31 ELEKGYQFDGW-----EISGPEGKQDAG-----YVINLSKDTFFIKPVPFKIE 73
Db 200 ELAVDYNFEWYGGAQVKVCLCGAVACSGFLGAKSRGQFASPYVILIEDTY---VWEDGDD 256
Qy 74 KGEENKETPDVSKKK--DNQVNVHSQLNESHKEDLQREH-----SQKSDSTKD- 122
Db 257 RYSVDKIEFYVDSAEDELTSBPSSKNGESNTBEKXKDISTENHLESTALNIQOQSDSTPTP 316
Qy 123 -----VTATVLDDKNNISKSSTTN 141
Db 317 MEEDVVTVETVKTETSEDMKLLSQN 340

RESULT 32
EB9883
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (stra
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: EB9883
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui
ma, A.; Mizutani-Di, Y.; Kobayashi, N.; Sawano, R.; Kaito, C.; Seki
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaishi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A99758; MUID:213111952; PMID:11418146

```

A:Accession: E89883  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <KUR>  
A:Cross-references: UNIPROT:Q99UK5; UNIPARC:UPI00000CAAE3; GB:BA000018; PID:g13700929; E89883  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0976

Query Match 11.9%; Score 89; DB 2; Length 645;  
Best Local Similarity 21.9%; Pred. No. 25;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;  
QY 6 TIONKEMSTIVSEEDFLPVYKGLB-KGYQF-----DGWEISGPEGKK-----50  
DB 355 SVNNESMMDTFVKH-----PIKTMGLNGKTKVMETNDDYWKDFMVEGQRVTRISKDA 409  
QY 51 -----DAGYVINL-SKDTPIKPFVKKIEKKEE 78  
DB 410 KNNRTIIPVYEGKTLDAIVKVHVHTIDYDQYHVRIVDKAEATKANTDKSNKKEOOD 469  
QY 79 NKPTFDV-----SKKDNQPNVHNSQLNESHKREDIQ-----REHSQKSDSTKQVT-ATVL 128  
DB 470 NSAKKEATPATPSKPTPSPEKESQKDSQKDNKQLPSVEKENDASSESGDKTPTATKP 529  
QY 129 DKNNTSSKSTNNPNK 144  
DB 530 TKGVESSSTT--PTK 543

RESULT 33  
T33009  
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
A:Accession: T33009  
R:Gentles, S.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z21815  
A:Accession: T33009  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1888 <GEN>  
A:Cross-references: UNIPROT:O14207; UNIPARC:UPI000013AR2F; EMBL:Z98531; PIDN:CAB11064.1;  
A:Experimental source: strain 972h-; cosmid c6B12  
C:Genetics:  
A:Gene: SPDB:SPAC6B12.02c  
A:Map position: 1  
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 11.9%; Score 89; DB 2; Length 1888;  
Best Local Similarity 23.9%; Pred. No. 82;  
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;  
QY 14 SSTIVSEEDF-----ILPVY---KGELEKGYQPDGWEISGFE-----GKDAQ 53  
DB 390 SSSLISENPFQUNVAANAVSTIPVTRTKTKKKH-RFYVEVEKLPDLILLESYGKKAPK 448  
QY 54 YVINLSKDTFIKPFVKKIEKKEENKPTFDVSKKDNQPNVHNSQLNESHKREDIQREH 113  
DB 449 FLRVFARSSSHIP--KMIRRKQMDSKKYFSPDKESDRQVIDQVLSDWYSGRHELVAQSH 506  
QY 114 SOKSDS-TKQVATVLDKN-----NISKSTNN 141  
DB 507 SYKQSPDSKSGVGNIFSVNSKKHVNINAKTAANN 541

RESULT 34  
T33068  
hypothetical protein C35E7.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T33068

R:Graves, T.; McDonald, R.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid C35E7.  
A:Reference number: Z21278  
A:Accession: T33068  
A:Status: preliminary;  
A:Molecule type: DNA  
A:Residues: 1-301 <GRA>  
A:Cross-references: UNIPROT:O61765; UNIPARC:UPI0000074979; EMBL:AF067216; PIDN:AAC17524.  
A:Experimental source: strain Bristol N2; clone C35E7  
C:Genetics:  
A:Gene: CESP:C35E7.9  
A:Map position: 1  
A:Introns: 30/3; 193/1; 236/2

Query Match 11.8%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;  
QY 35 GYQFDGWEISGPEGKK-----DAGYVINLSKDTFIKPFVKKI-----EE 73  
DB 10 GTTAGILAGCGGKKKKGKSSSTASAAAPKADSKMPPVENVKSKEKKEPKKEEP 69  
QY 74 KKEENKPTFDVSKKDNQPNVHNSQLNESHKREDIQREHSQKSDSTKD 122  
DB 70 KKEEKEKSKSEKDKDK-----EEAKKEDDKDEKDEKDKD 112

RESULT 35  
A71683  
hypothetical protein RP278 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
A:Accession: A71683  
R:Andersson, S.G.B.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alemark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71683  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-371 <AND>  
A:Cross-references: UNIPROT:Q9ZDP9; UNIPARC:UPI000000C1161; GB:AJ235269; NID  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: RP278

Query Match 11.8%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 15;  
Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;  
QY 11 KEMSTIVSEEDFLPVYKGELEKGYQPDGWEISGFEKK-DAGYVINLSKDTFIKPFVK 69  
DB 192 EQLNKTITNEB-----FRKNLEQ-----EKKKIEGKKTQASEALNKK----LKPIYK 234  
QY 70 KIEEKKEE-----ENKPTFDVSKKKNPNQV---HSOLNESHKREDIQREHSQKSDSTK 121  
DB 235 QMDSBELFKLAILFOY-----AQANIDKHALYAKQYQTKIENDPNYKELEKIQ 286  
QY 122 DVTATVLDKNNISSKSTT-----NNFN 143  
DB 287 E-----IVSKIEYSKSKNTDIINNPN 309

RESULT 36  
G88436  
protein T04A8.i3 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
A:Accession: G88436  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology



A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Accession: G88436  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-762 <STO>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI000017A5C4; GB:chr\_III; PIDN:CAA84732.1;  
C:Genetics:  
A:Gene: T04A8.13  
A:Map position: 3

Query Match 11.8%; Score 88.5; DB 2; Length 762;  
Best Local Similarity 25.2%; Pred. No. 33;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 19 SEEDFILPVYKGELEKGYQFGWEISGPEGK-KDAGYVINLSKDTFIPKVPKKEKKEE 77  
DB 14 SGKQIMKFGY--DKKEGLGMQKEIVGDDKKDKARKRKRLQDEFAR--LKQDEEKQKE 69  
QY 78 ENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKS 137  
DB 70 EAEKKEKKEKKEEKKEDGHEKKEDKKEDKKENENDEKSKSDKKESKEDKKEKTK 129  
QY 138 TTNNPNK 144  
DB 130 TEDNEGK 136

RESULT 37  
T23435  
hypothetical protein T04A8.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PIDN:CAA84732.2;  
A:Experimental source: clone T04A8  
C:Genetics:  
A:Gene: CESP:T04A8.13  
A:Map position: 3  
A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

Query Match 11.8%; Score 88.5; DB 2; Length 791;  
Best Local Similarity 25.2%; Pred. No. 34;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 19 SEEDFILPVYKGELEKGYQFGWEISGPEGK-KDAGYVINLSKDTFIPKVPKKEKKEE 77  
DB 14 SGKQIMKFGY--DKKEGLGMQKEIVGDDKKDKARKRKRLQDEFAR--LKQDEEKQKE 69  
QY 78 ENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKS 137  
DB 70 EAEKKEKKEKKEEKKEDGHEKKEDKKEDKKENENDEKSKSDKKESKEDKKEKTK 129  
QY 138 TTNNPNK 144  
DB 130 TEDNEGK 136

RESULT 38  
T23451  
hypothetical protein K08E3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23451  
R:McMurray, A.

submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23451  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-276 <WIL>  
A:Cross-references: UNIPROT:Q9KUT1; UNIPARC:UPI000007D88; EMBL:Z81568; PIDN:CAB04590.1  
A:Experimental source: clone K08E3  
C:Genetics:  
A:Gene: CESP:K08E3.2  
A:Map position: 3  
A:Introns: 81/3; 102/3; 169/1; 211/2

Query Match 11.8%; Score 88; DB 2; Length 276;  
Best Local Similarity 31.5%; Pred. No. 11;  
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 3;  
QY 58 LSKDTFIPKVPKKEKKEE--NKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQ 115  
DB 60 LKQDEVEKKKKGKEEKEEKKEDKKEKKDD-----SKEDKKDEHEK 107  
QY 116 KSDSTKDVTTATVLDKNNISSKSTTNNPNK 144  
DB 108 KK-----TAEKKEKKEKKEKKNK 128

RESULT 39  
G96796  
hypothetical protein P28016.18 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96796  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, K.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96796  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1280 <STO>  
A:Cross-references: UNIPROT:Q9SRD2; UNIPARC:UPI000000A3260; GB:AE005173; NID:G6143896; F

Query Match 11.8%; Score 88; DB 2; Length 1280;  
Best Local Similarity 23.7%; Pred. No. 63;  
Matches 32; Conservative 24; Mismatches 61; Indels 18; Gaps 3;  
QY 8 QNGKEMSTTVSEBDFILPVYKGELEKGYQFGWEISGPEGKDGAGYVINLSKDTFIPKVP 67  
DB 110 KGGGSVFALLDDDD-----EKEDNESDGDKDDPVIISFTGKKJASKKGGKGSFAASA 164  
QY 68 F-----KKTEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKS 117  
DB 165 FDALGSDDDDTTEVEHDEEESPTTSGKKKSS---KSSKKNNTSFTADLLDEEGTDA 221  
QY 118 DSTKDVTTATVLDKNN 132  
DB 222 SNRRDENTTEDEES 236

RESULT 40  
A41859  
19A-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae  
C:Species: Haemophilus influenzae

A:Variety: strain HK715  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41859  
R:Poulsen, K.; Reinholdt, J.; Kilian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty  
A:Reference number: A41859; MUID:92234949; PMID:1373717  
A:Accession: A41859  
A:Status: Preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1702 <POU>  
A:Cross-references: UNIPROT:P45384; UNIPARC:UPI000012D3F0; GB:M87489; NID:g148906; PIDN:  
A:Experimental source: strain HK715  
A:Note: sequence extracted from NCBI backbone (NCBIP:97282)  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 11.7%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 95;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;  
QY 56 INLSKDTFIKPVFKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLOREEHS- 114  
DB 1296 INTGSATAITETAESKDKPQTETAATEDASQHKANTVADNSVANNSESEPKGRRRSI 1355  
QY 115 --QKSDSTKQVTATVLDKNNISSKSTNNPNK 144  
DB 1356 SQPQTSABETTAETDDETTIADNSKRSKPNR 1387

Search completed: April 24, 2006, 15:01:13  
Job time : 16.1409 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 11.2752 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTIIONGKMSSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225428 segs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA\_New.\*
- 1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 4: /SIDSS/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 5: /SIDSS/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
  - 8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	14.7	746	6	US-10-793-626-652
2	108	14.4	354	7	US-11-189-817-2
3	105	14.0	651	7	US-11-128-660-1
4	89	11.9	501	6	US-10-485-517-381
5	89	11.9	645	6	US-10-485-517-244
6	87	11.6	140	7	US-11-096-568A-4771
7	86	11.5	443	7	US-11-188-298-1015
8	86	11.5	700	7	US-11-196-475-74
9	85.5	11.4	886	7	US-11-087-099-11456
10	85	11.4	472	6	US-10-793-626-658
11	85	11.4	700	7	US-11-196-475-66
12	85	11.4	708	7	US-11-196-475-76
13	85	11.4	720	6	US-10-793-626-2058
14	84.5	11.3	1694	7	US-11-052-554A-83
15	83.5	11.2	493	7	US-11-096-568A-3070
16	83.5	11.2	493	7	US-11-096-568A-3071
17	83.5	11.2	510	7	US-11-096-568A-3069
18	82	11.0	439	7	US-11-188-298-15964
19	82	11.0	439	7	US-11-188-298-16606
20	81.5	10.9	8746	7	US-11-098-686-10232
21	81	10.8	785	6	US-10-793-626-264
22	80.5	10.8	1036	7	US-11-096-568A-28315
23	80.5	10.8	1070	7	US-11-096-568A-28314
24	80.5	10.8	1276	7	US-11-096-568A-28313
25	79.5	10.6	943	6	US-10-475-204-34

26	79	10.6	568	6	US-10-793-626-2482	Sequence 2482, Ap
27	79	10.6	693	7	US-11-196-475-68	Sequence 68, Appl
28	79	10.6	732	7	US-11-124-368A-248	Sequence 248, App
29	79	10.6	803	7	US-11-124-368A-241	Sequence 241, App
30	79	10.6	803	7	US-11-124-368A-242	Sequence 242, App
31	79	10.6	1145	6	US-10-793-626-1432	Sequence 1432, Ap
32	78.5	10.5	258	7	US-11-098-686-10475	Sequence 10475, A
33	78.5	10.5	305	7	US-11-096-568A-17853	Sequence 17853, A
34	78.5	10.5	327	7	US-11-096-568A-17852	Sequence 17852, A
35	78.5	10.5	386	7	US-11-096-568A-17851	Sequence 17851, A
36	78.5	10.5	1155	6	US-10-793-626-1780	Sequence 1780, Ap
37	78	10.4	191	6	US-10-860-601-5	Sequence 5, Appli
38	77.5	10.4	405	6	US-10-485-517-239	Sequence 239, App
39	77.5	10.4	627	6	US-10-873-528-191	Sequence 191, App
40	77	10.3	513	7	US-11-185-924-16	Sequence 16, Appl
41	77	10.3	636	6	US-10-485-517-170	Sequence 170, Appl
42	77	10.3	663	7	US-11-196-475-70	Sequence 70, Appl
43	77	10.3	663	7	US-11-196-475-78	Sequence 78, Appl
44	77	10.3	693	7	US-11-196-475-72	Sequence 72, Appl
45	76.5	10.2	439	7	US-11-188-298-17915	Sequence 17915, A

ALIGNMENTS

RESULT 1  
US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	14.7%	Score 110;	DB 6;	Length 746;
Best Local Similarity	27.0%	Pred. No. 0.09;		
Matches	43;	Conservative 22;	Mismatches 52;	Indels 42; Gaps 8;
Qy	5	VTIIONGKMSSTIVSEDPILPVYK-----GLEKGYQPDG---EISGFE-----G 48		
Db	596	ITIGNKQIKQOSKSGTKVLPHSKVLMTDGELTMP-DMTGWTKEVDLAFEDLTCLKVS 654		
Qy	49	KKDAGYVYN--LSKDTFIPKVPFKIEKKEBENKPTFVS-----KKONPQVNSQLNES 102		
Db	655	TKGNQFVTVNQISIKGQIILK-----NKKIEVLSAEDTDDQKTDSDSDN 701		
Qy	103	HRKEDLQREHSQKSDSTKQVTVTLDRKNISSKSTNN 141		
Db	702	KSKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 734		

RESULT 2  
US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMIDIUM FALCIPARUM





```
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2003-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11456
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Bacillus cereus ATCC 14579
US-11-087-099-11456

Query Match      11.4%; Score 85.5; DB 7; Length 886;
Best Local Similarity 25.4%; Pred. No. 14;
Matches 46; Conservative 23; Mismatches 69; Indels 43; Gaps 9;

Qy  2 RVTVIQNGKEMSSSTIVSEEDFILPVY--KGE-----LEKGYOPDGHETSG- 45
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  300 RISIDIQRQKIPLEASVDEETKIPVYNLKGEPLOQNLQSEGATLNG--FIKWNSPGE 357
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  46 --PEGKDGAGVYNLSKDTFKIPVFKIEEKEEENKPTFDVSKK--DNPQVNHSQLNE 101
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  358 KIYEFALDTN--STESKIRFNGTIVQNIWEKQESQNVILDKTLOOHINKENLGRVNA 415
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  102 SHRKEDL----QREHSQKSDTKDVTN-----TVLDK-----NNISSKSTNNP 142
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  416 PITKEDLLQIKLEILKEGNEIKDTGLBYWNTNLEKLTLEGVGLKNTEPISNLKQLNNV 475
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  143 N 143
Db  476 N 476

RESULT 10
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match      11.4%; Score 85; DB 6; Length 472;
Best Local Similarity 32.6%; Pred. No. 7.1;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy  71 IEKKEEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQ-----REHSQKSDS--- 119
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  3 MEENKNQPNKE--NWSNKODNA-----THLNDSHRNEDLELPENKNAQRERRRIDNQSK 56
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  120 TKDVTAT-----VLDKRNISKSTNNPNK 144
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  57 EKDATSTQSQLETKPMDKFLDNHKS--HNQNK 86
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 11
US-11-196-475-66
; Sequence 66, Application US/11196475
; Publication No. US20050271682A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Query Match      11.4%; Score 85; DB 7; Length 700;
Best Local Similarity 24.4%; Pred. No. 11;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

Qy  3 VTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQPPGWEISGPEGKDGAVYNLSK-- 60
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  258 ITETIENLRDLEKATDEE-----HKGIES--QVDAKKQKEELDKA---INLDKQA 306
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  61 -----DTFKPVFKIEEKEEENKP-TFDVSKKDNQPNV-----HS 97
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  307 QKLDSADNLDVQRNVTREKIQEDINEINEKNLPKPGDVSSPKVDQLQIKESLEDLQE 366
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy  98 QLNES---HKEDLQREHSQSD-----STKDVATATVLDKRNISKSTNNPNK 144
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  367 QLKETGDNQKREIEKQIRIKSDEKULKSKDKKASKDGKALDLDLR-ELNSKASSKEKSK 425
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 12
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 708
```

; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi  
US-11-196-475-76

```

Query Match      11.4%; Score 85; DB 7; Length 708;
Best Local Similarity 27.6%; Pred. No. 12;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

Qy 3 VVTVIQNGKEMSSITVISEDPIILPVYKGELEKGYQFDGWEISGPEGKKDAGYVNLAK-- 60
Db 258 ITETIENLRDLEKATDEE-----HKKEETES--QVDAKKKKEBELDKCA---INLDRQA 306

Qy 61 -----DTPIKPVFKKI BEKGEENRP-TFDVSKKDNQPVNHSQNLNESHKK 105
Db 307 QKLDSAEINDLDVQRDVTREKIQEDINEINEKNLPPKPGDVSSPKVDQKL---QIKES--L 361

Qy 106 EDLQRE-----EHSQKSDSTKDVATVLDKNNISSK 136
Db 362 EDLQEQLEKAGDENQKREIKQIBIKKRDDELLAKK 397

```

```

RESULT 13
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

```

[illegible]

```

RESULT 14
US-11-052-554A-83
; Sequence 83, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052.554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589, 227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763

```

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-11-052-554A-83

```

```

Query Match      11.3%; Score 84.5; DB 7; Length 1694;
Best Local Similarity 26.8%; Pred.No.36;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 56 INSKDTFTIKPVFKIEKKKEENKCTFDVSKKQNPQVNHSQLNSHRRKEDLQREHHS- 114
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1288 INTGSATATETABKSDKQQTAASTEDASQKANTVADNSVANNSESDPKSRRRRI 1347

Qy 115 --QKSDSTKDTATVLDRKNINSKSTTNPNK 144

Db 1348 SQPQETASBETTAASTDETTIADNSKRSKPNR 1379

```

```

RESULT 15
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fr
; TITLE OF INVENTION: therapy
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

```

Query Match	11.2%	Score 83.5;	DB 7;	Length 493;
Best Local Similarity	25.7%;	Pred. No. 10;		
Matches	43;	Conservative 21;	Mismatches 62;	Indels 41; Gaps 8;

  

Qy	6	TIQNGKEM--SSTIVSEE-----	DFLLPVYKGELEKGYQPGWREISGPEGKKDAG	53
Db	80	TQSSGKQTADANTIVTEEEKKPGKVPKPKIKTPVSK---	KQETADSNKTETVTSDDKDEG	136
Qy	54	YVI-----	NLSKDTFIKVPFKCIEE--KQBEENKPTFVSKKONPQVNH	97
Db	137	NVAVAOQDDTQSTGTGTANADTTVTPEVKTKGVKKQSKTPT---	SEKRDN--TADS	191
Qy	98	QLMNESHKEDLQREHSQKSDSTKQVTAATVLDKNTISSTKSTTNPNK		144
Db	192	SKTETKSDKODKKEER-----	VTGEKSGAKTKLKASDKDQVNTNKGK	233

```

RESULT 16
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3071
; LENGTH: 493

```



```
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match
Best Local Similarity 11.2%; Score 83.5; DB 7; Length 493;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy 6 TIONKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 53
Db 80 TQSSGQTADANTIVTEKKPKGVVKKIKTPVSK---KKDETADSNKTETLSKKDEG 136
Qy 54 YVI-----NLSKDTPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNH 97
Db 137 NVVAQAQDDTOSTGQTANADTTVPEVKTKGVVKKQSKTPT---SEKRDN--TADS 191
Qy 98 QLNESHKEDLQREHSQKSDSKDVTATVLDKNNISSKSTNNPNK 144
Db 192 SKTETSKDDKKEER-----VTGKSGAKTDKLKASDKDVTNVKKG 233

RESULT 17
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592FUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc_feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match
Best Local Similarity 11.2%; Score 83.5; DB 7; Length 510;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy 6 TIONKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 53
Db 97 TQSSGQTADANTIVTEKKPKGVVKKIKTPVSK---KKDETADSNKTETLSKKDEG 153
Qy 54 YVI-----NLSKDTPIKVPFKIEE--KKEENKPTFDVSKKKNPQVNH 97
Db 154 NVVAQAQDDTOSTGQTANADTTVPEVKTKGVVKKQSKTPT---SEKRDN--TADS 208
Qy 98 QLNESHKEDLQREHSQKSDSKDVTATVLDKNNISSKSTNNPNK 144
Db 209 SKTETSKDDKKEER-----VTGKSGAKTDKLKASDKDVTNVKKG 250

RESULT 18
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US2006007552A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
```

```
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,97E
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964

Query Match
Best Local Similarity 11.0%; Score 82; DB 7; Length 439;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 35 GYQFDGWEISGFEGKKDAGVYVNLKDTPT-----KP-----VFK 69
Db 45 GISPDGSSVPGFQGIEDSLVFKADPTTYVEVPMDNVARVGYFYKONKPYGADPRGLK 104
Qy 70 KIEKKEEN-----KPTFDVSKKKN----- 91
Db 105 RALELEKEGKAYIGPEPEFYLPKNGTWELEIPDVGGYFDILTLDKARDIRREIAEYM 164
Qy 92 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 133
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 19
US-11-188-298-16606
; Sequence 16606, Application US/11188298
; Publication No. US2006007552A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 16606
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus woesei
US-11-188-298-16606

Query Match
Best Local Similarity 11.0%; Score 82; DB 7; Length 439;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 35 GYQFDGWEISGFEGKKDAGVYVNLKDTPT-----KP-----VFK 69
Db 45 GISPDGSSVPGFQGIEDSLVFKADPTTYVEVPMDNVARVGYFYKONKPYGADPRGLK 104
Qy 70 KIEKKEEN-----KPTFDVSKKKN----- 91
Db 105 RALELEKEGKAYIGPEPEFYLPKNGTWELEIPDVGGYFDILTLDKARDIRREIAEYM 164
Qy 92 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 133
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 20
US-11-098-686-10232
; Sequence 10232, Application US/11098685
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
```

; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: PatSeq for Windows Version 4.0  
; SEQ ID NO 10232  
; LENGTH: 8746  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10232

Query Match 10.9%; Score 81.5; DB 7; Length 8746;  
Best Local Similarity 25.9%; Pred. No. 4.6e+02;  
Matches 38; Conservative 24; Mismatches 60; Indels 25; Gaps 7;  
QY 12 EMSSTIVSEEDFIL--PVYKGLKGYQPDGWE--ISGFEGKDGAGYVINLSKDTFFIKPV 67  
Db 3358 EGEDVAEEDLILLWDLDNEAEGTKEHAEVKGVEGVEFDG-----ISED--KP- 3410  
QY 68 FKKEBEKEENKPTFDVSKKONPQVNHSQLNESHKEDL-----QREHSQKS 117  
Db 3411 --KDDKEQEQATLGDSSGRTIEESQPPQEEBEKKENSPGSGNESPSPQEEESYDE 3468  
QY 118 DSTKDVATVLDKNNISSKSTNNPNK 144  
Db 3469 TSSVVTSPLSLINEV--KQTEDEKSAK 3493

RESULT 21  
US-10-793-626-264  
; Sequence 264, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P034800S  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 264  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-264

Query Match 10.8%; Score 81; DB 6; Length 785;  
Best Local Similarity 24.7%; Pred. No. 29;  
Matches 37; Conservative 27; Mismatches 66; Indels 20; Gaps 7;  
QY 4 TWTIONGKMS-STIVSEEDFILPVYKGLKGYQPDGWEISGFEGKDGAGY-VINLSKD 61  
Db 638 TITEENGIVCITVTYGEDLMMVTNAGVI---IFLDVHDISQ-NGRAAQGVRLMKLGDG 693  
QY 62 TFIKPVFKKIEKKKEEN-----KPTFDVSKKONPQVNHSQLNESHKEDLQRE 112  
Db 694 QFVSTVAKVNEEDDNEENADRAQSTTTTETADVEEVD---DQTFPGNAIHTEGDAEWES 749  
QY 113 -HSQKSDSTKDVATVLDKNNISSKSTNN 141  
Db 750 VEPENDRIDIRQDFMDRVNEDIESA DN 779

RESULT 22  
US-11-096-568A-28315  
; Sequence 28315, Application US/11096568A

; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28315  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURES:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1036)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712010  
US-11-096-568A-28315

Query Match 10.8%; Score 80.5; DB 7; Length 1036;  
Best Local Similarity 24.6%; Pred. No. 44;  
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 5;  
QY 9 NGKEMSTIVSEEDFILPVYKGL-EGYQPDGWEISGFEGKDGAGYVINLSKDTF---- 63  
Db 516 NEQTPAKTIIVKKK-IIRKVAKRKVAEIDNKMVG--DSKKDGDSDKKVMGVKKSSDGS 572  
QY 64 --IKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREH---SOKSD 118  
Db 573 VEMKPTAESLEDVKDENASKTVDVKQETGSPDTKKKGASSSSKKDTGTGEDKKAERKNN 632  
QY 119 STKDVATVLDKNNISSK 136  
Db 633 SETMSEKGIKIDRNNNTDEK 650

RESULT 23

US-11-096-568A-28314  
; Sequence 28314, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28314  
; LENGTH: 1070  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURES:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1070)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712009  
US-11-096-568A-28314

Query Match 10.8%; Score 80.5; DB 7; Length 1070;  
Best Local Similarity 24.6%; Pred. No. 46;  
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 5;  
QY 9 NGKEMSTIVSEEDFILPVYKGL-EGYQPDGWEISGFEGKDGAGYVINLSKDTF---- 63  
Db 550 NEQTPAKTIIVKKK-IIRKVAKRKVAEIDNKMVG--DSKKDGDSDKKVMGVKKSSDGS 606  
QY 64 --IKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREH---SOKSD 118  
Db 607 VEMKPTAESLEDVKDENASKTVDVKQETGSPDTKKKGASSSSKKDTGTGEDKKAERKNN 666  
QY 119 STKDVATVLDKNNISSK 136  
Db 667 SETMSEKGIKIDRNNNTDEK 684





```
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1432
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1432
```

```
Query Match          10.6%; Score 79; DB 6; Length 1145;
Best Local Similarity 24.8%; Pred. No. 67;
Matches 33; Conservative 22; Mismatches 44; Indels 34; Gaps 6;

Qy  9 NGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKSDTFIKPVF 68
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  408 NGHAVNEKLNGEYDVVPSLISEL---YQTQ-TSISN-----TYVL-----TFSDVI 451
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  69 KKEBKKEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQKSDSTKDVATVL 128
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  452 KALNKKIENESTPLFEA-----VNHVQVNELSSDENEDRYEYDR-----YI 493
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  129 DQNISSKSTTN 141
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  494 ELNFKDSLTSN 506
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

## RESULT 32

```
US-11-098-686-10475
; Sequence 10475, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10475
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10475
```

```
Query Match          10.5%; Score 78.5; DB 7; Length 258;
Best Local Similarity 22.0%; Pred. No. 12;
Matches 20; Conservative 26; Mismatches 30; Indels 15; Gaps 3;

Qy  58 LSKDTFIKPVFKKIEKKEENKPT-FDVSKKK-----DNPQVNHSQLNESHKEDLQ 109
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  81 IEKQPIKPNLRFWHELKQDNPNPLYNKEKKIHNLSQIEDTNLQHSQIPSSHPQDLK 140
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  110 RE-----EHSQKSDSTKDVATVLDKNI 133
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  141 OKQDISSETKNIQRNINTEQVQQQTTSQNDL 171
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

## RESULT 33

```
US-11-096-568A-17853
; Sequence 17853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17853
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(305)
; OTHER INFORMATION: Ceres Seq. ID no. 12361233
US-11-096-568A-17853
```

```
Query Match          10.5%; Score 78.5; DB 7; Length 305;
Best Local Similarity 19.4%; Pred. No. 15;
Matches 37; Conservative 22; Mismatches 67; Indels 65; Gaps 4;

Qy  3 VVTVIQKEMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKSDT 62
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  43 VDCSLSGDAGSSKKKAEKSSFRPAK-----ETPSLEDSNEKKKTKQASNQH 90
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  63 FIKPVFKKIEKKEENKP-----TFDVSKKKDNQPNVNHSQLNESHKEDLQRE 111
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  91 SVK---KDIIEESNESVKPORVGSTPSYGFSPKCDERAERKREFYSKLEKIHAQELEKS 147
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  112 EHSQKSDSTKDVATVLDKN-----N 132
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  148 NLQAKSKETEAEELKMLRKSINPKATPMPSFYKEPPPKVELKKIPITTRARSPKLGSKN 207
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  133 ISSKSTTNPN 143
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  208 TSSGGTEGNPN 218
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

## RESULT 34

```
US-11-096-568A-17852
; Sequence 17852, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17852
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(327)
; OTHER INFORMATION: Ceres Seq. ID no. 12361232
US-11-096-568A-17852
```

```
Query Match          10.5%; Score 78.5; DB 7; Length 327;
Best Local Similarity 19.4%; Pred. No. 16;
Matches 37; Conservative 22; Mismatches 67; Indels 65; Gaps 4;

Qy  3 VVTVIQKEMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKSDT 62
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  65 VDCSLSGDAGSSKKKAEKSSFRPAK-----ETPSLEDSNEKKKTKQASNQH 112
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  63 FIKPVFKKIEKKEENKP-----TFDVSKKKDNQPNVNHSQLNESHKEDLQRE 111
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  113 SVK---KDIIEESNESVKPORVGSTPSYGFSPKCDERAERKREFYSKLEKIHAQELEKS 169
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy  112 EHSQKSDSTKDVATVLDKN-----N 132
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 170 NIQAQSKETEEAEELKMLKSLNFKATPMPSEFYKBPBPPKVELKCIPTTRARSPKLGSRKN 229  
QY 133 ISSKSTTNPN 143  
Db 230 TSSGGTEGNPN 240  
RESULT 35  
US-11-096-568A-17851  
; Sequence 17851, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 17851  
; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(386)  
; OTHER INFORMATION: Ceres Seq. ID no. 12361231  
US-11-096-568A-17851  
Query Match 10.5%; Score 78.5; DB 7; Length 386;  
Best Local Similarity 19.4%; Pred. No. 20;  
Matches 37; Conservative 22; Mismatches 67; Indels 65; Gaps 4;  
QY 3 VTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQDGEWISGSPGKKGAGVIVNLKSDT 62  
Db 124 VDCSLNGDAGSKKKAKKSSFRPAK-----ETPSLSDSNEKKKTQKASQNH 171  
QY 63 FIKPVFKIEEKKEBENKPD-----TPDVSKKKONPOVNHISQLNESHKEDLQRE 111  
Db 172 SVK---KQIEESNESVQPVGVSTPSYGFSPKCDERAERKEFPYKLEEKIHAQLEKRS 228  
QY 112 EHSQKSDSTKVATVLDKN-----N 132  
Db 229 NIQAQSKETEEAEELKMLKSLNFKATPMPSEFYKBPBPPKVELKCIPTTRARSPKLGSRKN 288  
QY 133 ISSKSTTNPN 143  
Db 289 TSSGGTEGNPN 299  
RESULT 36  
US-10-793-626-1780  
; Sequence 1780, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1780  
; LENGTH: 1155  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1780

Query Match 10.5%; Score 78.5; DB 6; Length 1155;  
Best Local Similarity 27.3%; Pred. No. 74;  
Matches 24; Conservative 12; Mismatches 27; Indels 25; Gaps 2;  
QY 73 EKKEENKPTFDVSKKKDN-----PQVNHSQLNESHKEDLQREHSQK 116  
Db 100 EVKVEAPTTSDSVKPKANEAVVTNESTPKTKTTPATVNEESIATPKTSTTQODSTSKN 159  
QY 117 SDSTKDVATVLDKNKNISSKSTTNPNK 144  
Db 160 NPSLKD-----NLNSSSTTSKSK 178  
RESULT 37  
US-10-860-601-5  
; Sequence 5, Application US/10860601  
; Publication No. US20050273871A1  
; GENERAL INFORMATION:  
; APPLICANT: Asano, Yoshihiro  
; APPLICANT: Takashima, Seiji  
; APPLICANT: Kitakaze, Masafumi  
; TITLE OF INVENTION: Method for Diagnosing Arrhythmogenic Right Ventricular Dysplasia  
; FILE REFERENCE: 2144.0210000  
; CURRENT APPLICATION NUMBER: US/10/860,601  
; CURRENT FILING DATE: 2004-06-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: RVAP27  
US-10-860-601-5  
Query Match 10.4%; Score 78; DB 6; Length 191;  
Best Local Similarity 23.9%; Pred. No. 9.6;  
Matches 27; Conservative 29; Mismatches 37; Indels 20; Gaps 4;  
QY 14 SSTIVSEDFILP-----VYKGELEKGYQPDGW--HISGPEGKKGAGVIVNLKSDT 65  
Db 11 SSSSEDEEYVYKVLDRRVYKGVYLLKWKGFSEHNTWPEK-----NLDCPELIS 64  
QY 66 PVFKIKIEKKEENKPTFDVSKKKONPOVNHISQLNESHKEDLQREHSQKSD 118  
Db 65 EPMKKYKKKEGEN-----NKPREKSESNNKKNFNSNSADDDIKSKKKKEQSN 111  
RESULT 38  
US-10-485-517-239  
; Sequence 239, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosyntex Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629WO  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 239  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-239

Query Match 10.4%; Score 77.5; DB 6; Length 405;  
Best Local Similarity 29.7%; Pred. No. 26;  
Matches 22; Conservative 16; Mismatches 33; Indels 3; Gaps 2;  
  
QY 69 KKIEEKKEENKPTFDVSKKONPQVNHSQLN-ESHRKEDLQREHSHQKSDSTKDVATV 127  
DB 51 KAAESTNKLNEATTASDQSDKVDVMDQLNQEDNTRKNDQKENVSSQGNETTNGNKL 110  
  
QY 128 LDKNNISSKSTNN 141  
DB 111 IEKESV--QSTTGN 122

RESULT 39  
US-10-873-528-191  
; Sequence 191, Application US/10873528  
; Publication No. US20050276814A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/10/873,528  
; CURRENT FILING DATE: 2004-06-23  
; PRIOR APPLICATION NUMBER: US/09/769,787  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 191  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-873-528-191

Query Match 10.4%; Score 77.5; DB 6; Length 627;  
Best Local Similarity 23.4%; Pred. No. 44;  
Matches 30; Conservative 26; Mismatches 59; Indels 13; Gaps 4;  
  
QY 30 GLEKGY--QPDGWEISGEGKKDAGYVNLKOTFIKVPVKKIEKKEENKPT----- 82  
DB 497 GAMKQGHKKANDWPFYKTDGSRVAGWIKDKDKYFLKENGQLLVNGKTPGTYVDSSGA 556  
  
QY 83 --PDVSKKON--PQVNHSQLNESH--KEDLQREHSHQKSDSTKDVATVLDKNNISSK 136  
DB 557 WLVDVSIKSAIKITTSHEIKESKEVVKDLLENKETSQHSVTFNFTSQDLTSSTSSS 616  
  
QY 137 STNNPNK 144  
DB 617 ETSVNKSE 624

RESULT 40  
US-11-185-924-16  
; Sequence 16, Application US/11185924  
; Publication No. US20060078945A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher et al., Larry  
; TITLE OF INVENTION: Complex Formed by Small Integrin-Binding Ligand,  
; TITLE OF INVENTION: N-Linked Glycoproteins (SIBLINGS) and Factor H  
; FILE REFERENCE: 4239-61301-02  
; CURRENT APPLICATION NUMBER: US/11/185,924  
; CURRENT FILING DATE: 2005-07-19  
; PRIOR APPLICATION NUMBER: 09/958,617  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: PCT/US00/09349  
; PRIOR FILING DATE: 2000-04-09  
; PRIOR APPLICATION NUMBER: 60/128,468

; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-185-924-16

Query Match 10.3%; Score 77; DB 7; Length 513;  
Best Local Similarity 23.0%; Pred. No. 38;  
Matches 32; Conservative 30; Mismatches 57; Indels 20; Gaps 6;  
  
QY 8 QNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFFIKPV 67  
DB 350 ENSSEQEEVVSESRGDNP-----DPTTSYVEDQEDSDSSE--EDSSHTLSHKS----- 397  
  
QY 68 FKIEEKKEENKPTFDVSKKK-DNPQVNHSQLN-----SHRKEDLQREHSHQKSDS-T 120  
DB 398 -ESREEQADSESSESINPSESPESPEDENSSSQEGLOSHSSSAESQSESHSEEDSDS 456  
  
QY 121 KDVATVLDKNNISSKSTT 139  
DB 457 QDSRSKEDSNSTESKSSS 475

Search completed: April 24, 2006, 15:44:48  
Job time : 12.2752 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 72.1611 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTVTIQNGKMSSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:  
1: /cgm2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgm2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgm2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgm2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgm2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgm2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	773	4	US-10-067-385-8
2	748	100.0	2119	3	US-09-769-744A-28
3	748	100.0	2140	4	US-10-282-122A-73670
4	748	100.0	2140	5	US-10-472-928-1180
5	745	99.6	637	3	US-10-617-320-3169
6	615	82.2	117	3	US-09-765-272-68
7	615	82.2	117	6	US-11-106-649-68
8	110	14.7	778	4	US-10-724-972A-5663
9	108	14.4	188	5	US-10-691-672A-7
10	106.5	14.2	470	5	US-10-739-930-6262
11	105	14.0	647	5	US-10-691-672A-3
12	104	13.9	665	3	US-09-820-843A-107
13	103.5	13.8	169	5	US-10-691-672A-2
14	102.5	13.7	775	4	US-10-282-122A-70721
15	101.5	13.6	564	6	US-11-097-143-12723
16	99.5	13.3	707	4	US-10-282-122A-52942
17	95	12.7	973	5	US-10-732-923-18783
18	93.5	12.5	898	4	US-10-425-115-205148
19	93	12.4	869	4	US-10-437-963-12282
20	93	12.4	948	5	US-10-732-923-4286
21	93	12.4	1529	5	US-10-732-923-8762
22	93	12.4	3127	5	US-10-732-923-22588
23	92.5	12.4	540	5	US-10-732-923-22820
24	92.5	12.4	2468	4	US-10-755-889-615
25	92.5	12.4	2468	5	US-10-489-740-216
26	92.5	12.4	2519	5	US-10-450-763-46995
27	92	12.3	1005	4	US-10-437-963-187665

28	91.5	12.2	903	4	US-10-282-122A-52328	Sequence 52328, A
29	91	12.0	2060	4	US-10-381-596A-2	Sequence 2, Appl
30	90	12.0	932	4	US-10-282-122A-52510	Sequence 52510, A
31	89.5	12.0	511	4	US-10-289-762-509	Sequence 509, App
32	89.5	12.0	1408	6	US-11-097-143-2904	Sequence 2904, App
33	89	11.9	519	4	US-10-437-963-160737	Sequence 160737, A
34	89	11.9	645	4	US-10-282-122A-70234	Sequence 70234, A
35	89	11.9	645	5	US-10-470-048B-414	Sequence 414, App
36	89	11.9	654	4	US-10-172-502-10	Sequence 10, Appl
37	89	11.9	654	6	US-11-020-509-10	Sequence 10, Appl
38	88.5	11.8	281	6	US-11-097-143-21276	Sequence 21276, A
39	88	11.8	225	4	US-10-032-585-7829	Sequence 7829, Ap
40	88	11.8	645	5	US-10-470-048B-142	Sequence 142, App
41	88	11.8	815	5	US-10-496-905-23	Sequence 23, Appl
42	88	11.8	1980	5	US-10-482-834A-144	Sequence 144, App
43	87.5	11.7	441	5	US-10-510-812-14	Sequence 14, Appl
44	87.5	11.7	717	5	US-10-732-923-23071	Sequence 23071, A
45	87.5	11.7	1196	4	US-10-282-122A-52737	Sequence 52737, A

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 748;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.8e-58;	Mismatches 0;	Indels 0; Gaps 0;
Matches 144;	Conservative 0;			
Qy	1	HRVTVTIQNGKMSSTIVSEDFILPVYKGLERKGYQPDGWEISGFEGKDGAGYVNI-SK 60		
Db	630	HRVTVTIQNGKMSSTIVSEDFILPVYKGLERKGYQPDGWEISGFEGKDGAGYVNI-SK 689		
Qy	61	DTFIPKVPFKIEEKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 120		
Db	690	DTFIPKVPFKIEEKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 749		
Qy	121	KDVTATVLDKNNISSKSTNNPNK 144		
Db	750	KDVTATVLDKNNISSKSTNNPNK 773		

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins



Publication No. US20050136404A1  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
FOR DIAGN  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Denke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...637  
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
US-10-617-320-3169  
Query Match 99.6%; Score 745; DB 5; Length 637;  
Best Local Similarity 99.3%; Pred. No. 1.1e-57;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 HRVTVTIQNGKEMSTTVSEEDFILPVYKGBLEKGYQPDGWEISGFEKGDAGYVINLSK 60  
Db 460 HRVTVTIQNGKEMSTTVSEEDFILPVYKGBLEKGYQPDGWEISGFEKGDAGYVINLSK 519  
Qy 61 DTFIKVPVKIEEKKBEENKPTFDVSKKDPQVNHSQLNESHKEDLQREHSQKSDST 120  
Db 520 DTFIKVPVKIEEKKBEENKPTFDVSKKDPQVNHSQLNESHKEDLQREHSQKSDST 579  
Qy 121 KDVTATVLDKNNISSKSTNNPNK 144  
Db 580 KDVTATVLDKNNISSKSTNNPNK 603  
RESULT 6  
US-09-765-272-68  
Sequence 68, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-272-68  
Query Match 82.2%; Score 615; DB 3; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.3e-47;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 28 YKGBLEKGYQPDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKKBEENKPTFDVSK 87  
Db 1 YKGBLEKGYQPDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKKBEENKPTFDVSK 60  
Qy 88 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNNISSKSTNNPNK 144  
Db 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNNISSKSTNNPNK 117  
RESULT 7  
US-11-106-649-68  
Sequence 68, Application US/11106649  
Publication No. US20050181439A1  
GENERAL INFORMATION:  
APPLICANT: Choi et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
FILE REFERENCE: PB340P2C3D1  
CURRENT APPLICATION NUMBER: US/11/106,649  
CURRENT FILING DATE: 2005-04-15  
PRIOR APPLICATION NUMBER: US 09/765,271  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: US 09/536,784  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: US 08/961,083  
PRIOR FILING DATE: 1997-10-30  
PRIOR APPLICATION NUMBER: US 60/029,960  
PRIOR FILING DATE: 1996-10-31  
NUMBER OF SEQ ID NOS: 454  
SOFTWARE: Patent in version 3.3  
SEQ ID NO 68

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      82.2%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.3e-47;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YKGELEKGYQPGWISGPEGKQDGYVINLSKDTFIKVPFKIIEKKKEENKPTFDVSK 87
Db 1 YKGELEKGYQPGWISGPEGKQDGYVINLSKDTFIKVPFKIIEKKKEENKPTFDVSK 60

QY 98 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 61 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S.epidermidis
US-10-724-972A-5663

Query Match      14.7%; Score 110; DB 4; Length 778;
Best Local Similarity 27.0%; Pred. No. 0.44;
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFGW---EISGFE-----G 48
Db 628 ITIGNGKQIKQOSVKSXGTVKVLPHSKVLMWTDGELTWP-DMTGWTXEDVLAFEDLTIKIVS 686

QY 49 KKDAGYVIN--LISKDTFIKVPFKIIEKKKEENKPTFDVS-----KKDNPQVNHSQLNES 102
Db 687 TKGNGFVNTQSIKQIILK-----NKKDIEVLSAEDTDDQKTDSDSDN 733

QY 103 HRKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 141
Db 734 KSKDKDAEDHSHNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 9
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
```

```
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match      14.4%; Score 108; DB 5; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.12;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWISGPF--EGKQDAG-----YVINLS 59
Db 15 VLKAKEASS-----YDYIL-----GWFGGGVPEHKKEENMLSHLYVSSKD 55

QY 60 KDTFIKVPFKIIEKKKEE-----ENKPTFDVSKKKNPQVNHSQLNESHKKE 106
Db 56 KENISKENDVDLDEKKEAEETEEELKKNKEETEEISEDEEEEEEKKEEEDKCK 115

QY 107 DLQREHSHQKSDSTKDVATVLDKNNISSKSTNN 141
Db 116 EQEKEQSNENDQKQMEA-----QNLISKNNNN 145

RESULT 10
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6262
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270_1.p
US-10-739-930-6262

Query Match      14.2%; Score 106.5; DB 5; Length 470;
Best Local Similarity 20.3%; Pred. No. 0.49;
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWISGFE-----GKQDA 52
Db 84 NRVTDTVQNNNGSK-----YVQDLARRIRYDE-EATGSSQSAQRIDHPNQNV 131

QY 53 GYVINLSKDTFIKVPFKIIEKKKEENKPTFDVSKKKN-----PQ 93
Db 132 GITEKAFENSPIEETSHRVDDNKNINQKNFTAAKSSSENNAVSRVSGADHKKRAEYMGKPM 191

QY 94 VNHSQLNE-----SHRKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 192 ENRDQVRQTESAEKSHRKNVTKSEKPRDQGVKKTEAKDKDRNKEKKEETESINK 248

RESULT 11
US-10-691-672A-3
; Sequence 3, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
```

```
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-3

Query Match      14.0%; Score 105; DB 5; Length 647;
Best Local Similarity 22.7%; Pred. No. 0.98;
Matches 42; Conservative 35; Mismatches 52; Indels 56; Gaps 9;

QY 1 HRVTVTIQNGEM-----SSTIVSEEDFILPVYKGELEK-----GYQ 37
DB 432 HEETVSQSNPEKANDGNVQSNNELNEFV-----ESEKSEHARSKAKEASSYD 485
QY 38 FD-GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKIEKKKEENKPTFDVSKK 89
DB 486 YILGWEFGGVPYKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEHEELE 544
QY 90 DNPQVNHSQLN-----ESHKRBQLQREHSQKSDSKDTKVATVLDKNNISSK 136
DB 545 KNEBETESEISEDEREEBEKEENKKEQKEQESQNNENDQKKMEA-----QNLISK 599
QY 137 STTN 141
DB 600 NQNN 604

RESULT 12
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: g1|3845248
US-09-820-843A-107

Query Match      13.9%; Score 104; DB 3; Length 665;
Best Local Similarity 25.9%; Pred. No. 1.2;
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

QY 1 HRVT-VTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFEGKDGAGYVIN 57
DB 138 HRQNELNLQSGK-----NEQDI-----NKNKGGKQ-----DTSNSNAENKKD----- 174
QY 58 LSKDTFIKPVFKIEKKE-----EENKPTFD-----VSKKQNDPQVNHSQLNESH 103

; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-2

Query Match      13.8%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.26;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 11 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 63
DB 1 KEASS-----YDYIL-----GWFGGVPYKKEENMLSHLYVSSKDKENI 41
QY 64 IKPVFKIEKKKEENKPTFDVSKKQNDPQVNHSQLN-----ESHKRBQLQ 110
DB 42 SKENDVDLDE-KEEAEETEHEELEKKEEETEHEISEDEEEEEEEKEEKKQGEK 100
QY 111 EHSQKSDSKDTKVATVLDKNNISSKSTTN 141
DB 101 EQSNENDQKKMEA-----QNLISKQNNN 126

RESULT 13
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match      13.8%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.26;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 11 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 63
DB 1 KEASS-----YDYIL-----GWFGGVPYKKEENMLSHLYVSSKDKENI 41
QY 64 IKPVFKIEKKKEENKPTFDVSKKQNDPQVNHSQLN-----ESHKRBQLQ 110
DB 42 SKENDVDLDE-KEEAEETEHEELEKKEEETEHEISEDEEEEEEEKEEKKQGEK 100
QY 111 EHSQKSDSKDTKVATVLDKNNISSKSTTN 141
DB 101 EQSNENDQKKMEA-----QNLISKQNNN 126

RESULT 14
US-10-282-122A-70721
; Sequence 70721, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70721
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70721

Query Match
Best Local Similarity 13.7%; Score 102.5; DB 4; Length 775;
Matches 38; Conservative 26; Mismatches 61; Indels 29; Gaps 7;

QY 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW-----EISGFE-----G 48
Db 625 ITIGNGKIQKQSVKSGTKVLPKSHKVMWLTGDELTPM-DMTGWTKEVDLAFEDLTKIKVS 683
QY 49 KKDAGYVYN--LSKDTFFKPVK-----KIEKKKEENKPTPDVSKKKONPOVNHISQ 98
Db 684 TKNGFVNTQISKGQIIKNKDKIEVLSAEDTDDQKTDDESDSKSKDKVDDEDSN 743
QY 99 LNESHRKEDLQREHSQKSDSKDVTATVLDKNN 132
Db 744 ASSSKNEKSNADSKNDSDDSTNETSGS--ERN 775

RESULT 15
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match
Best Local Similarity 13.6%; Score 101.5; DB 6; Length 564;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 21 EDFILPVYKGELEKGYQFDGW-----EISGFEKKDAGYVI-----NLSKDTFFIK 65
Db 78 EDLDTPLSESRSFK--VFDGWVDEHRDHDGHDVQEPSGEALDDHDEHDDHDEDEDEE 135
QY 66 PVFKKIEKKKEENKPT-----PDVSKKKONPOVNHISQKSHKEDLQREHSQKSDS 119
Db 136 PLTELEEELEEEETPEDEPADEFEDEENNA--GENITAEDAESEEEEDND 193
QY 120 TKDVTATVLDKNNISSKST 138
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 16
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC FEATURE
```

LOCATION: (6)..(6)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (18)..(18)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (29)..(29)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (37)..(37)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (43)..(43)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (54)..(54)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (84)..(84)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (86)..(86)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (359)..(359)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (385)..(385)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (388)..(388)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (396)..(396)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (400)..(400)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (402)..(402)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (404)..(404)  
OTHER INFORMATION: X=any amino acid  
US-10-282-122A-52942

Query Match 13.3%; Score 99.5; DB 4; Length 707;  
Best Local Similarity 25.3%; Pred. No. 3.4;  
Matches 37; Conservative 26; Mismatches 60; Indels 23; Gaps 4;  
Qy 18 VSEDFILPVYKGELEKGYQFDGWEISGPEKKGADGVINLSKDTPIKVPVKLEEK--- 74  
Db 484 ISIEDDAEGVKEBIDSNNDIGDVV---EDKDTTDEYDSNKEDIIPEPNKSKGKAKL 540  
Qy 75 -----KBEEN---KPTFDVSKKDNQVNHSQLNESHKEDLQREHSHOKSD 118  
Db 541 FGPIKONFEVQBEENLNDISPDILDKPVENNQVSKSEIEQNELKE-IKQEPFSQIIE 599  
Qy 119 STKDVATVLDKNNISSKSTTNNPNK 144

Db 600 EERSVKIEKFINNNLDEKVSNNESK 625  
RESULT 17  
US-10-732-923-18783  
; Sequence 18783, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 18783  
; LENGTH: 973  
; TYPE: PRT  
; ORGANISM: Plasmodium yoelii yoelii  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(973)  
; OTHER INFORMATION: unsure at all xaa locations  
US-10-732-923-18783

Query Match 12.7%; Score 95; DB 5; Length 973;  
Best Local Similarity 21.8%; Pred. No. 12;  
Matches 31; Conservative 37; Mismatches 68; Indels 6; Gaps 3;  
Qy 1 HRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISGPEKKGADGVINLSK 60  
Db 601 YNINMNDQKGECNSTYKT---LIQHRSGSKESGKNKFIHTKINNISRHARGVYSLSK 656  
Qy 61 DTPIKVPVKLEKBEENKPTFDVSKKDNQVNHSQLNESHKEDL-QREHSHOKSDS 119  
Db 657 DNNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDN 716  
Qy 120 TKDVTATVLDKNNISSKSTTNN 141  
Db 717 T-FVTATSNKKNNDINKNSND 737

RESULT 18  
US-10-425-115-205148  
; Sequence 205148, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 205148  
; LENGTH: 898  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_118684C.1.pap  
US-10-425-115-205148

Query Match 12.5%; Score 93.5; DB 4; Length 898;  
Best Local Similarity 23.5%; Pred. No. 15;  
Matches 36; Conservative 32; Mismatches 62; Indels 23; Gaps 6;  
Qy 8 QNGKEMSTIVSEDF-----ILPVYKGELEK-----GYQFDGWEISGPEKKGADGVINLSK 52  
Db 8 QNGKEMSTIVSEDF-----ILPVYKGELEK-----GYQFDGWEISGPEKKGADGVINLSK 52





```
QY 11 KEMSTTVSEEDFILPVYKGELEKGYQPDGWEISG---PBGKADAGYVNLSDTFIKPV 67
DB 2000 KKEITVTSDDMFTSPVNIKEYNEQERKKEIVGNLSYDKTKIPFPFIKTFKEGRIRK- 2058
QY 68 FKIEKEKEENK-----PTF-----DVSKKONQOV 94
DB 2059 -KKIEKEKEKENNNPFLYNDYSSSPKYGDNENNFVIKYIRERKDPQKKFDHPNF 2117
QY 95 NISQL-----NESHK---EDLQREHSQKSDTKD-VYATVLDKNNISSTNNP 142
DB 2118 NPSKFLNPNPKNNKNNKNNKNNRNEYPNTSSKDGVSYNFLSDFSSDNEYSSD 2177
QY 143 NK 144
DB 2178 NE 2179

RESULT 23
US-10-732-923-22820
; Sequence 22820, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22820
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-22820

Query Match 12.4%; Score 92.5; DB 5; Length 540;
Best Local Similarity 24.8%; Pred. No. 10;
Matches 32; Conservative 25; Mismatches 55; Indels 17; Gaps 4;

QY 10 KEMSTTVSEEDFILPVYKGELEKGYQPDGWEISGFGKADAGYVNLSDTFIKPVK 69
DB 47 GVEKASFKESDFPADLKESEK-----ALSDLKSLKEAIVDN-----TLTKTKK 94
QY 70 K---IEEKKKEENKPTFDVSKKKONPQVNHSQLNESHKEDLQREHSQKSDTKDVTAT 126
DB 95 BSSPMKEKEEVVKEAEVKEKKK--EAAERKVEEKKSEAVVTEAPKAEVAVVTEE 152
QY 127 VLDKNNIS 135
DB 153 IIPKEVTT 161

RESULT 24
US-10-755-889-615
; Sequence 615, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NP-kB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 615
; LENGTH: 2468
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-755-889-615

Query Match 12.4%; Score 92.5; DB 4; Length 2468;
Best Local Similarity 31.0%; Pred. No. 65;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 58 LSKDTFTKVPVKIEKKKEENKPTFDVSKKKONPQVNHSQLNESHKEDLQRE-----E 112
DB 638 VKKETVKVP-----EDKKEKEKPKCKGVAKKEDKTP-----KKEEKPKCKEVEKCKEIK 689
QY 113 HSQKSDSTKOV 123
DB 690 KEEKKPKCKEV 700

RESULT 25
US-10-489-740-216
; Sequence 216, Application US/10489740
; Publication No. US20050112574A1
; GENERAL INFORMATION:
; APPLICANT: Biomimics Limited
; TITLE OF INVENTION: P9
; FILE REFERENCE: Angiogenesis PCT
; CURRENT APPLICATION NUMBER: US/10/489,740
; CURRENT FILING DATE: 2004-03-15
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 216
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-489-740-216

Query Match 12.4%; Score 92.5; DB 5; Length 2468;
Best Local Similarity 31.0%; Pred. No. 65;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 58 LSKDTFTKVPVKIEKKKEENKPTFDVSKKKONPQVNHSQLNESHKEDLQRE-----E 112
DB 638 VKKETVKVP-----EDKKEKEKPKCKGVAKKEDKTP-----KKEEKPKCKEVEKCKEIK 689
QY 113 HSQKSDSTKOV 123
DB 690 KEEKKPKCKEV 700

RESULT 26
US-10-450-763-46995
; Sequence 46995, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIPI/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 46995
; LENGTH: 2519
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: DOMAIN
; LOCATION: (1040)..(1091)
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
```

```
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1919)..(2122)
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by Pfam,
; OTHER INFORMATION: accession name MAP1B_neuraxin, E-value=1.9e-59, Pfam score of 190
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2519)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-46995

Query Match          12.4%; Score 92.5; DB 5; Length 2519;
Best Local Similarity 31.0%; Pred. No. 66;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 58 LSKOTFIKPVFKIEKKEENKPTFDVSKGNQVNVHSQLNESHKEDLQRE-----E 112
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
689 VKKETKVP-----EDKKEKEKPKKEVAKKEDKTPI---KKEEKPKKEEVKKEVKEIK 740
QY 113 HSKSDSTKOV 123
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
741 KEEKPKPKKEV 751

RESULT 27
US-10-437-963-187665
; Sequence 187665, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazul, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187665
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84346C.1.pep
US-10-437-963-187665

Query Match          12.3%; Score 92; DB 4; Length 1005;
Best Local Similarity 26.4%; Pred. No. 24;
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;

QY 45 GREGKDGAGYVNLKSDTFIKVPFKIEKKEENKPTFDVSKK-KDNQV--NHSQVNE 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 GSEKEMSGKNIKSIKETGTGQSKELQKESKRSKTKDKSKKNKDKMTQVPTNABEFHK 94
QY 102 SH-----RKE-----DLQREHSOKSDSTKQVTTATVLDKNNISSKST 138
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
95 EYTKVIRKESDSSIEQVIGTSSIQEMETNEQVSKDTSKDMTQVPAANAGIRKEYT 154
QY 139 T 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
155 T 155

RESULT 28
US-10-282-122A-52328
; Sequence 52328, Application US/10282122A
; Publication No. US20040029129A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52328
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52328

Query Match          12.2%; Score 91.5; DB 4; Length 903;
Best Local Similarity 25.7%; Pred. No. 23;
Matches 38; Conservative 25; Mismatches 50; Indels 35; Gaps 8;

QY 5 VTIONGKEMSTIVSEBDFILPVYKGELEKGYQDQGWHEISGPEGKKDAGYVNLKSDTFI 64
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 INVELKEEKSKQIIEKHNEL-----KNBKSISPKESIKL---KKEKD----- 229
QY 65 KPVPKTEEKKEENKPTFDVS-----KKDNQVNVHSQLNESHKEDLQREHSOKSD 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 KQVFAHNESNEENKQISKVNVELKNERSKQLPKIN-VELKEENKQSI--KEHNELRE 286
QY 119 STKQVTTATVLDKNNISSKSTNN--PNK 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 BTK----KCLPKVNIELKEETKKQVPNK 310

RESULT 29
US-10-381-596A-2
; Sequence 2, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
; APPLICANT: Biostapro AB
; TITLE OF INVENTION: von Willebrand factor-binding proteins from
; TITLE OF INVENTION: Staphylococci
; FILE REFERENCE: 110059600
; CURRENT APPLICATION NUMBER: US/10/381,596A
```





Db 470 NSAKKEATPATPSKPTSPVEKESQKDSQKDNKQLPSVEKENDASSESGDKTPTATKP 529  
QY 129 DKNNISKSSTNNPNK 144  
Db 530 TKGEVSSSTT--PTK 543

## RESULT 35

US-10-470-048B-414  
; Sequence 414, Application US/10470048B  
; Publication No. US20050037444A1  
; GENERAL INFORMATION:  
; APPLICANT: MEINKE ET AL.  
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF  
; FILE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN  
; FILE REFERENCE: SONN:035US  
; CURRENT APPLICATION NUMBER: US/10/470,048B  
; CURRENT FILING DATE: 2003-07-25  
; NUMBER OF SEQ ID NOS: 603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 414  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-470-048B-414

Query Match 11.9%; Score 89; DB 5; Length 645;  
Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50  
Db 355 SVNNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQVRVTISKDA 409  
QY 51 -----DAGVIVNL-SKDTFIKVPFKKIBKKKEE 78  
Db 410 KNNTRTIIPPVVEGKTLDAIVKVHVKTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 469  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----REHSQKSDSTKDVT-ATVL 128  
Db 470 NSAKKEATPATPSKPTSPVEKESQKDSQKDNKQLPSVEKENDASSESGDKTPTATKP 529  
QY 129 DKNNISKSSTNNPNK 144  
Db 530 TKGEVSSSTT--PTK 543

## RESULT 36

US-10-172-502-10  
; Sequence 10, Application US/10172502  
; Publication No. US20030185833A1  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .  
; FILE REFERENCE: P07263US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/172,502  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-172-502-10

Query Match 11.9%; Score 89; DB 4; Length 654;  
Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;  
QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50

Db 364 SVNNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQVRVTISKDA 418  
QY 51 -----DAGVIVNL-SKDTFIKVPFKKIBKKKEE 78  
Db 419 KNNTRTIIPPVVEGKTLDAIVKVHVKTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 478  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----REHSQKSDSTKDVT-ATVL 128  
Db 479 NSAKKEATPATPSKPTSPVEKESQKDSQKDNKQLPSVEKENDASSESGDKTPTATKP 538  
QY 129 DKNNISKSSTNNPNK 144  
Db 539 TKGEVSSSTT--PTK 552

## RESULT 37

US-11-020-509-10  
; Sequence 10, Application US/11020509  
; Publication No. US20050106648A1  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .  
; FILE REFERENCE: P07263US02/BAS  
; CURRENT APPLICATION NUMBER: US/11/020,509  
; CURRENT FILING DATE: 2004-12-27  
; PRIOR APPLICATION NUMBER: US 10/172,502  
; PRIOR FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-11-020-509-10

Query Match 11.9%; Score 89; DB 6; Length 654;

Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50  
Db 364 SVNNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQVRVTISKDA 418  
QY 51 -----DAGVIVNL-SKDTFIKVPFKKIBKKKEE 78  
Db 419 KNNTRTIIPPVVEGKTLDAIVKVHVKTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 478  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----REHSQKSDSTKDVT-ATVL 128  
Db 479 NSAKKEATPATPSKPTSPVEKESQKDSQKDNKQLPSVEKENDASSESGDKTPTATKP 538  
QY 129 DKNNISKSSTNNPNK 144  
Db 539 TKGEVSSSTT--PTK 552

## RESULT 38

US-11-097-143-21276  
; Sequence 21276, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832

```
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21276
; LENGTH: 281
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-21276
```

```
Query Match      11.8%; Score 88.5; DB 6; Length 281;
Best Local Similarity 22.1%; Pred. No. 10;
Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;

QY  9 NKGKMSSTIVSEEDFILPVYKGELEKGYQDFDGEISGPEGKDGAGYVI-----NLS 59
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 NFEETNMNTTQODLLKNY-SLLDK-----DNEGATTSKELGNVIRALGRQPNESIA 188
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  60 KDTFIKVPFKLEE-KCEEENKPTFDVSKKDNQPNVHSQLNE-----SHRKEDI-QREEH 113
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 KEFCNVILURKHDHTKEELKDAFRVFDKENNGYISTTELRAVFWALGEKLEDELEEM 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 114 SQKSDSTKDVATVLDKNNI 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 IREYDLDDQNHINFEESNNM 268
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 39
US-10-032-585-7829
; Sequence 7829, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7829
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7829
```

```
Query Match      11.8%; Score 88; DB 4; Length 225;
Best Local Similarity 28.0%; Pred. No. 8.8;
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY  36 YQFDGWEISGFGKGDAGYVINLSKDTFIKVPFKK-----IEBKGEENKPTDVSQKK-- 89
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  80 YDDDDDEFGFESSNGAAKELNLSQALKEWKQRDLIEERELKNSKKKEEIEIAKAS 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  90 --DNPQVNSQLNESHKEDLQREH--SQKSDSTKDVATVLDKNN 132
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 TIDDFYNTNKRDNHQKILSEQEKFIKRDDFLK--RGTLDWRVN 184
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 40

```
US-10-470-048B-142
; Sequence 142, Application US/1047004fB
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:0350S
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-142
```

```
Query Match      11.8%; Score 88; DB 5; Length 645;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY  6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 SVENNESMMDTFVXH-----PIKTGMLNGKKYMYMETTNDYWKDFWVEGQRTISKDA 409
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  51 -----DAGYVINL-SKOTFIKVPFKKIEBKKEE 78
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 KNNTRTIIPYVEGKTLDAIVKHVKTIDYDGYHVRIVDKAFTKANTDKSNKEQOD 469
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  79 NKPTFDV-----SKKDNQPNVHSQLNESHKEDLQ-----REEHSQKSDSTKDV--ATVL 128
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 NSAKKEATPATPSKPTSPVEKESQKQDSQKODNKQLPSVEKENDASSESGKGYTLATKP 529
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 129 DKNNISSKSTTNNPNK 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 530 TKGEVESSTT--PTK 543
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: April 24, 2006, 15:43:34  
Job time : 74.1611 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 22.3893 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTQNGKEMSTIVSE.....ATVLDKNKISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*  
1: /cgm2\_6/ptodata/1/iaa/5 COMB.pep:\*  
2: /cgm2\_6/ptodata/1/iaa/6 COMB.pep:\*  
3: /cgm2\_6/ptodata/1/iaa/H COMB.pep:\*  
4: /cgm2\_6/ptodata/1/iaa/PCTUS COMB.pep:\*  
5: /cgm2\_6/ptodata/1/iaa/RE COMB.pep:\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	100.0	773	2	US-09-590-991-8
2	745	99.6	637	2	US-09-107-433-3169
3	745	99.6	2138	2	US-09-583-110-5274
4	615	82.2	117	2	US-08-961-083-68
5	615	82.2	117	2	US-09-536-784-68
6	615	82.2	117	2	US-09-765-271-68
7	615	82.2	117	2	US-09-765-272A-68
8	110	14.7	746	2	US-09-710-279-652
9	110	14.7	778	2	US-09-134-001C-3868
10	101.5	13.6	347	2	US-09-248-796A-16224
11	95	12.7	348	2	US-09-538-092-1316
12	92.5	12.4	2468	2	US-09-976-594-726
13	92.5	12.4	2468	2	US-09-538-092-1135
14	92.5	12.4	2522	2	US-09-949-016-10237
15	89.5	12.0	511	2	US-09-198-452A-509
16	89.5	12.0	511	2	US-09-438-185A-475
17	89	11.9	654	2	US-10-172-502-10
18	88	11.8	280	2	US-09-248-796A-17646
19	87.5	11.7	1702	2	US-08-296-791-5
20	87.5	11.7	1702	2	US-09-839-986-5
21	87.5	11.7	1702	2	US-10-080-505-5
22	87.5	11.7	1702	2	US-10-645-655-5
23	87.5	11.7	1702	4	PCT-US95-10661A-5
24	87	11.6	902	2	US-09-134-001C-5157
25	86	11.5	172	2	US-09-248-796A-21065
26	86	11.5	243	2	US-09-248-796A-20306
27	86	11.5	700	2	US-08-235-836C-74

28	86	11.5	758	2	US-09-949-016-8288	Sequence 8288, Ap
29	86	11.5	1315	2	US-09-200-850E-5	Sequence 5, Appli
30	86	11.5	1989	2	US-09-949-016-10076	Sequence 10076, A
31	85.5	11.4	109	2	US-09-248-796A-24668	Sequence 24668, A
32	85.5	11.4	402	2	US-09-464-483-4	Sequence 4, Appli
33	85.5	11.4	402	2	US-09-414-664-4	Sequence 4, Appli
34	85.5	11.4	529	2	US-09-464-483-2	Sequence 2, Appli
35	85.5	11.4	529	2	US-09-414-664-2	Sequence 2, Appli
36	85	11.4	465	2	US-09-134-001C-3856	Sequence 3856, App
37	85	11.4	472	2	US-09-710-279-658	Sequence 658, App
38	85	11.4	700	1	US-07-720-589-2	Sequence 2, Appli
39	85	11.4	700	1	US-08-785-190-2	Sequence 2, Appli
40	85	11.4	700	2	US-08-235-836C-66	Sequence 66, Appl
41	85	11.4	700	4	PCT-US92-05359-2	Sequence 2, Appli
42	85	11.4	708	2	US-08-235-836C-76	Sequence 76, Appl
43	85	11.4	720	2	US-09-710-279-2058	Sequence 2058, Ap
44	85	11.4	728	2	US-09-134-001C-4968	Sequence 4968, Ap
45	85	11.4	743	2	US-08-910-925-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score 748;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 5.9e-71;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	HRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK	60	
Db	630	HRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK	689	
Qy	61	DTFKPVPKLEKKKEBKPTDVSKKDNPQNHSQLNESHKRDLOREHSOKSDST	120	
Db	690	DTFKPVPKLEKKKEBKPTDVSKKDNPQNHSQLNESHKRDLOREHSOKSDST	749	
Qy	121	KDVTATVLDKNKISSKSTNNPNK	144	
Db	750	KDVTATVLDKNKISSKSTNNPNK	773	

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD/ROM ISO9660
/ COMPUTER: <Unknown>
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/ 085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 3169:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 637 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...637
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
/
/ US-09-107-433-3169
/
/
/ Query Match 99.6%; Score 745; DB 2; Length 637;
/ Best Local Similarity 99.3%; Pred. No. 9.5e-71;
/ Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 60
/ Db
/ 460 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 519
/
/ QY 61 DTFIKPVFKIEEKEENKPTFDVSKKONPVNHSQNLNESHKEDLQREHHSQKSDST 120
/ Db
/ 520 DTFIKPVFKIEEKEENKPTFDVSKKONPVNHSQNLNESHKEDLQREHHSQKSDST 579
/
/ QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
/ Db
/ 580 KDVTATVLDKNNISSKSTTNNPNK 603
/
/ RESULT 3
/ US-09-583-110-5274
/ Sequence 5274, Application US/09583110
/ Patent No. 6699703
/
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al.
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
/ TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
/ FILE REFERENCE: PATH00-07A
/ CURRENT APPLICATION NUMBER: US/09/583,110
/ CURRENT FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/107,433
/ PRIOR FILING DATE: 1998-06-30
/ PRIOR APPLICATION NUMBER: US 60/085,131
/ PRIOR FILING DATE: 1998-05-12
/ PRIOR APPLICATION NUMBER: US 60/051,553
/
```

```
/ PRIOR FILING DATE: 1997-07-02
/ NUMBER OF SEQ ID NOS: 5322
/ SEQ ID NO 5274
/ LENGTH: 2138
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ US-09-583-110-5274
/
/ Query Match 99.6%; Score 745; DB 2; Length 2138;
/ Best Local Similarity 99.3%; Pred. No. 4.9e-70;
/ Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 60
/ Db
/ 1961 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVINLSK 2020
/
/ QY 61 DTFIKPVFKIEEKEENKPTFDVSKKONPVNHSQNLNESHKEDLQREHHSQKSDST 120
/ Db
/ 2021 DTFIKPVFKIEEKEENKPTFDVSKKONPVNHSQNLNESHKEDLQREHHSQKSDST 2080
/
/ QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
/ Db
/ 2081 KDVTATVLDKNNISSKSTTNNPNK 2104
/
/ RESULT 4
/ US-08-961-083-68
/ Sequence 68, Application US/08961083
/ Patent No. 6159469
/
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-961-083-68
/
/ Query Match 82.2%; Score 615; DB 2; Length 117;
/ Best Local Similarity 100.0%; Pred. No. 6.7e-58;
/ Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 28 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKIEEKEENKPTFDVSK 87
```

Db 1 YGGELEKGYQFDGWEISGFEKGKADGYVINLSKDTFKVPFKKIEKKEENKPTFDVSK 60  
QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 144  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/536,784  
; FILING DATE: 30-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-536-784-68

Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.7e-58;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YGGELEKGYQFDGWEISGFEKGKADGYVINLSKDTFKVPFKKIEKKEENKPTFDVSK 87  
Db 1 YGGELEKGYQFDGWEISGFEKGKADGYVINLSKDTFKVPFKKIEKKEENKPTFDVSK 60  
QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 144  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,271  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/536,784  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: OCT-30-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB340P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-271-68

Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.7e-58;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YGGELEKGYQFDGWEISGFEKGKADGYVINLSKDTFKVPFKKIEKKEENKPTFDVSK 87  
Db 1 YGGELEKGYQFDGWEISGFEKGKADGYVINLSKDTFKVPFKKIEKKEENKPTFDVSK 60  
QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 144  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPK 117

## RESULT 7

US-09-765-272A-68  
; Sequence 68, Application US/09765272A  
; Patent No. 6929930  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and  
; NUMBER OF SEQUENCES: 454  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: Dell Latitude C610  
; OPERATING SYSTEM: Windows 2000  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:



Db 260 PVFKTLQDLREB-----WRAEKBOANPKKBEENLNQKPVAKQKQKQPNSTKQKQKQKQ 314  
Qy 120 TKDVTATVLDKNNISSKSTNNPNK 144  
Db 315 TKKITPKTSKMLSGISTNLIINK 339

RESULT 11  
US-09-538-092-1316  
; Sequence 1316, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1316  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316

Query Match 12.7%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.068;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
Qy 29 KGELEKGFQDGEWISGEGKKDAGVINLSKDTPIKPVKKIIEKKBEENKPTP---DV 85  
Db 205 ESEGEK-----GTERKSKGKDS-----KXGKDSAIELQAVKADKKEDGKKDANKGDE 256  
Qy 86 SK--KKDNPVNHSQLN-----BSHRKEDLQREHSQKSDSTKD---VTATVLDKNNI 133  
Db 257 SKDAKKDAKEIKKGGKKKSPSTDSKDDVKKE---SKDATTQAKKVAKKOTKEKA 313  
Qy 134 SSK 136  
Db 314 DSK 316

RESULT 12  
US-09-976-594-726  
; Sequence 726, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 726  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1  
US-09-976-594-726  
Query Match 12.4%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 1.8;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 58 LSKDTPIKPVFKIIEKKBEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQRE-----E 112  
Db 638 VKKETKVKP-----EDKKEKEKEKPKCKEVAKKEDKTPI---KKEBKPKCKEYKCKEYKCKEIK 689  
Qy 113 HSQKSDSTKDV 123  
Db 690 KEEKCKPKCKEY 700

RESULT 13  
US-09-538-092-1135  
; Sequence 1135, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1135  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P46821  
US-09-538-092-1135

Query Match 12.4%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 1.8;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 58 LSKDTPIKPVFKIIEKKBEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQRE-----E 112  
Db 638 VKKETKVKP-----EDKKEKEKEKPKCKEVAKKEDKTPI---KKEBKPKCKEYKCKEYKCKEIK 689  
Qy 113 HSQKSDSTKDV 123  
Db 690 KEEKCKPKCKEY 700

RESULT 14  
US-09-949-016-10237  
; Sequence 10237, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USBS THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08



;; PRIOR APPLICATION NUMBER: US 60/074,725  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 17646  
;; LENGTH: 280  
;; TYPE: PRT  
;; ORGANISM: Candida albicans  
US-09-248-796A-17646

Query Match 11.7%; Score 88; DB 2; Length 280;  
Best Local Similarity 25.8%; Pred. No. 0.28; Mismatches 25; Indels 34; Gaps 9;  
Matches 39; Conservative 25; Mismatches 53; Indels 34; Gaps 9;

QY 4 TTVTNGKEMSTIVSBEEDFILPVYKLEKGYQPDGWEISGFEGKQDAGVYVNLKSDTF 63  
DB 77 TTVTKT--SVASTFCCKYDFNVFVSNLSLGF-----ELYSYANK-----KNSF 120  
QY 64 IKPVFKIE-EKKEENKPTFDVSKKONPQVNH-S-QLNESHKEDLQREHS----- 114  
DB 121 --PSPFHEIHSSEBK----YLKKHPQLQHHNLHHLHQRVPIKSHKVEGNRTIIN 174  
QY 115 --OKSDSTKDVATVLDKNNISSKSTNNPN 143  
DB 175 PQNLNDVYHINPTLLSSNG-STSTNNEN 204

RESULT 19  
US-08-296-791-5  
; Sequence 5, Application US/08296791  
; Patent No. 6245337  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-08-296-791-5

Query Match 11.7%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 3.6; Mismatches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 56 INLSKDTFIKPVFKIEKKEENKPTFDVSKKONPQVNH-S-QLNESHKEDLQREHS- 114  
DB 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPFKRRRSI 1355  
QY 115 --OKSDSTKDVATVLDKNNISSKSTNNPNK 144  
DB 1356 SQPQTSABETTAASTDTETIADNSKSKPNR 1387

RESULT 20  
US-09-839-996-5  
; Sequence 5, Application US/09839996  
; Patent No. 6642371  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839,996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-839-996-5

Query Match 11.7%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 3.6; Mismatches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 56 INLSKDTFIKPVFKIEKKEENKPTFDVSKKONPQVNH-S-QLNESHKEDLQREHS- 114  
DB 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPFKRRRSI 1355  
QY 115 --OKSDSTKDVATVLDKNNISSKSTNNPNK 144  
DB 1356 SQPQTSABETTAASTDTETIADNSKSKPNR 1387

RESULT 21  
US-10-080-505-5  
; Sequence 5, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.



```
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-080-505-5

Query Match      11.7%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKVPFKKIEKKEEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQETSAEETAASTDETTIADNSKRSKPNR 1387

RESULT 22
US-10-645-655-5
; Sequence 5, Application US/10645655
; Patent No. 6815182
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; PROTEIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/645,655
; FILING DATE: 20-Aug-2003
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 23-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5
```

```
Query Match      11.7%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKVPFKKIEKKEEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQETSAEETAASTDETTIADNSKRSKPNR 1387

RESULT 23
PCT-US95-10661A-5
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: PP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-5

Query Match      11.7%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKVPFKKIEKKEEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQETSAEETAASTDETTIADNSKRSKPNR 1387

RESULT 24
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
; Patent No. 6380370
```

[illegible]

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-836C-74

Query Match 11.5%; Score 86; DB 2; Length 700;
Best Local Similarity 23.2%; Pred. No. 1.6;
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;

QY 3 VTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD-----GWEISFGEGKK 50
Db :|||::: || :||: || :||: || :||: || :||: || :||: || :||: ||
258 ITETIENLRDQLEKATDSE-----HKCBIES--QVDAKKKQKEBLDKAIDLDRKAQQL 309
QY 51 D-AGVYNLSKDTFKPVFKIEKKKEENKPTFDVSKKDNQPVN-----HSQLN 100
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
310 DPAEDNLDIQDRTVREKLEQMINETNKEKNLPKPGDVSPKVKQLQIKESLEDLQEQIK 369
QY 101 ES----HRKEDLQREHSQKSD-----STKDVTTATVLDKNNISSKSTNNPNK 144
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
370 ETGDNQKRETEKQIEIKKSDKLLKSKDDKASKDGKALDLDLDR--ELNSKASSKEKSK 425

RESULT 28
US-09-949-016-8288
; Sequence 8288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8288
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8288

Query Match 11.5%; Score 86; DB 2; Length 758;
Best Local Similarity 29.1%; Pred. No. 1.8;
Matches 44; Conservative 13; Mismatches 52; Indels 42; Gaps 8;

QY 9 NGKMSSTIVSE--EDFILPVYKGELEKGYQF-----DGWE-----ISGFEGKK 50
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
410 HGKSDSPNVYTEKKEIAILRELTELKLTPEQQRSDLWERLYVEAKDQNGKQGTGKK 469
QY 51 DAG----YVINLSKDTFTKPVFKIEKKKEENKPTFDVSKKDNQPVNH-----SQLNE 101
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
470 KGRGSRHAKNKSKEFTLGSV-----KETFDAMKNSTKEFVRHHKKEIKQAKE 517
QY 102 SHRKEDLQREHSQKSD--STKDVTTATVLDK 130
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
518 A-VKENLKKFSDSVKSTFRHFKDTTKNIFDE 547

RESULT 29
US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.

```





```
RESULT 36
US-09-134-001C-3856
; Sequence 3856, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3856
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3856

Query Match 11.4%; Score 85; DB 2; Length 465;
Best Local Similarity 32.6%; Pred. No. 1.2;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 71 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQ-----REHSQKSDS--- 119
D 3 MEENKQPKB--NMSKODNA---THLNSHRNEDLELFRNKNARQRRRRIDNQS 56

QY 120 TKDVTAT-----VLDKNNISSKSTTNNPK 144
D 57 EKDATSQLETKPMDKFDNHS--HNQNK 86

RESULT 37
US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-658

Query Match 11.4%; Score 85; DB 2; Length 472;
Best Local Similarity 32.6%; Pred. No. 1.2;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 71 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQ-----REHSQKSDS--- 119
D 3 MEENKQPKB--NMSKODNA---THLNSHRNEDLELFRNKNARQRRRRIDNQS 56

QY 120 TKDVTAT-----VLDKNNISSKSTTNNPK 144
D 57 EKDATSQLETKPMDKFDNHS--HNQNK 86

RESULT 38
US-07-720-589-2
```

```
; Sequence 2, Application US/07720589
; Patent No. 5324630
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/720,589
; FILING DATE: 19910628
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-720-589-2

Query Match 11.4%; Score 85; DB 1; Length 700;
Best Local Similarity 24.4%; Pred. No. 2;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

QY 3 VVTIQGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGYVNLK-- 60
D 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKEELDKCA---INLDKAQ 306

QY 61 -----DTFIKPVFKIEKKEHENKP-TFDVSKKQNPQV-----HS 97
D 307 QKLSAEDNLDVQRTVREKIQEDINEKKNLPKPGDVSSPKVDKQLQIKESLEDLQE 366

QY 98 QLNES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTTNNPK 144
D 367 QKSTGDNQKREKQIEKKEDEKLLKSKDKDKKALDLDLR-ELNSKASKEKSK 425

RESULT 39
US-08-785-190-2
; Sequence 2, Application US/08785190
; Patent No. 5977339
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,190
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/720,589
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-785-190-2

Query Match 11.4%; Score 85; DB 1; Length 700;
Best Local Similarity 24.4%; Pred. No. 2;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

Qy 3 VVTVIQKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGVYNLSK-- 60
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKQELDKKA---INLDKQAQ 306
Qy 61 -----DTFIKPVFKIEKKKEENKP-TFDVSKKKDNQPQN-----HS 97
Db 307 QKLDSAEADNLDVQRNTVREKIQEDINEINKENLPGDGVSSPKVDKQLQIKESLEDLQE 366
Qy 98 QLNES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTNNPNK 144
Db 367 QLKETGDNQKREIEKQIEIKKSDKLLKSKDDKASKDGKALDLDLR-ELNSKASSKEKSK 425

RESULT 40
US-08-235-836C-66
; Sequence 66, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-235-836C-66

Query Match 11.4%; Score 85; DB 2; Length 700;
Best Local Similarity 24.4%; Pred. No. 2;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

Qy 3 VVTVIQKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGVYNLSK-- 60
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKQELDKKA---INLDKQAQ 306
Qy 61 -----DTFIKPVFKIEKKKEENKP-TFDVSKKKDNQPQN-----HS 97
Db 307 QKLDSAEADNLDVQRNTVREKIQEDINEINKENLPGDGVSSPKVDKQLQIKESLEDLQE 366
Qy 98 QLNES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTNNPNK 144
Db 367 QLKETGDNQKREIEKQIEIKKSDKLLKSKDDKASKDGKALDLDLR-ELNSKASSKEKSK 425

Search completed: April 24, 2006, 15:03:37
Job time : 23.3893 secs
```